ELECTRONICS

Also by
F. G. SPREADBURY

ELECTRIC DISCHARGE LIGHTING

A practical book on electric discharge lamps of most of the types available for lighting practice. There are many useful tables of data for the various lamps and curves showing the run-up characteristics. **15s.** net.

"*The diagrams and illustrations are clear and well reproduced and the typography is admirable.*" —ELECTRICAL REVIEW.

"*Clear and concise. Information that many readers will find useful and interesting.*"—NATURE.

AIRCRAFT ELECTRICAL ENGINEERING

A thoroughly up-to-date and comprehensive guide for designers, instructors and everyone concerned with electrical work in aircraft. **21s.** net.

"*Can be recommended to the student and the designer as containing much information of fundamental value.*"—ENGINEERING.

"*An excellent summary of the special features associated with design for aircraft requirements.*"—TECHNICAL JOURNAL.

PITMAN BOOKS

ELECTRONICS

BY

F. G. SPREADBURY, A.M.Inst.B.E.

LECTURER IN PHYSICS AND MATHEMATICS
WORKING MEN'S COLLEGE, LONDON
AUTHOR OF "AIRCRAFT ELECTRICAL ENGINEERING"
"ELECTRIC DISCHARGE LIGHTING"

OBSERVATORY,
LIBRARY.

LONDON
SIR ISAAC PITMAN & SONS, LTD.
1947

IIA Lib.,
01345

SIR ISAAC PITMAN & SONS, Ltd.
PITMAN HOUSE, PARKER STREET, KINGSWAY, LONDON, W.C.2
THE PITMAN PRESS, BATH
PITMAN HOUSE, LITTLE COLLINS STREET, MELBOURNE
UNITEERS BUILDING, RIVER VALLEY ROAD, SINGAPORE
27 BECKETTS BUILDINGS, PRESIDENT STREET, JOHANNESBURG

ASSOCIATED COMPANIES
PITMAN PUBLISHING CORPORATION
2 WEST 45TH STREET, NEW YORK
205 WEST MONROE STREET, CHICAGO
SIR ISAAC PITMAN & SONS (CANADA), Ltd.
(INCORPORATING THE COMMERCIAL TEXT BOOK COMPANY)
PITMAN HOUSE, 381–383 CHURCH STREET, TORONTO

BOOK
PRODUCTION
WAR ECONOMY
STANDARD

THIS BOOK IS PRODUCED IN
COMPLETE CONFORMITY WITH THE
AUTHORIZED ECONOMY STANDARDS

MADE IN GREAT BRITAIN AT THE PITMAN PRESS, BATH
D7—(T.289)

PREFACE

LIBRARY.

During the last decade of the nineteenth century there occurred a series of discoveries which revolutionized physical outlook and initiated what is usually termed modern physics. These discoveries were those of X-rays by Röntgen, radio-activity by Becquerel, and the electron by Thomson. Prior to these events it was thought by many physicists that the main discoveries of physics had been made and that the future would merely see their development. The previous physical discoveries are now generally referred to as classical physics.

The outstanding features of modern physics comprise the three phenomena mentioned, the quantum theory, the theory of relativity, and the atomic theory. In the hands of their initiators and many eminent followers, these discoveries and theories have undergone (and are still undergoing) striking developments, with the result that the growth of physical knowledge during the past fifty years probably exceeds that of any previous similar period.

As technical processes consist of applied physics, it is to be anticipated that the new regime in physics has been responsible for outstanding engineering developments. This is the case, the most notable example being doubtlessly in the field of electricity. Thus, the present tendency is for applied electricity to fall into two broad divisions, namely, electrodynamics and electronics. The first of these is based mainly on classical physics and the latter on modern physics.

As the term implies, electronics is based on the concept of the electron and its behaviour under specified conditions. Hence, electronics seeks to explain electrical phenomena, both pure and applied, in terms of electron behaviour. In particular it attempts to explain the facts of electrical conduction in metals and gases, the behaviour of charged particles (electrons and ions) in combinations of electrical and magnetic fields, electron focusing, and luminescence. Its field of inquiry on the technical side is principally concerned with gas-filled and vacuum valves, X-rays, static rectifiers, photo-electric cells, electric discharge lamps, cathode-ray tubes, besides a host of miscellaneous equipment such as the cyclotron, the electron microscope, etc.

The early chapters of the present book deal with what may be termed pure electronics and the later chapters with the principal applications. As a knowledge of the kinetic, quantum, and atomic theories is essential to the understanding of electronic matters, these theories are briefly dealt with in Chapter I. It is assumed that the reader is familiar with the principles of the main branches of classical physics and also that his mathematical standard includes the calculus.

The author is glad to acknowledge his indebtedness to the following firms who have kindly provided information and illustrations relative to their products: B.T.H. Co., Ltd.; G.E.C., Ltd.; Cinema-Television, Ltd.; Radio Corporation of America; Siemens Electric Lamps and Supplies, Ltd.; Philips Industrial; and the Edison Swan Electric Co., Ltd.

In addition, special thanks are due to the author's wife, who was responsible for the preparation of the manuscript and to whose assistance, both direct and indirect, any merit which this book may have is largely due.

F. G. SPREADBURY

LONDON, N.W.2, 1946

CONTENTS

PAGE

PREFACE v

CHAPTER I

ELECTRONICS AND THE FINE STRUCTURE OF MATTER—THE KINETIC, QUANTUM, AND ATOMIC THEORIES 1

Introduction—Origin of Electrons—Atoms and Molecules—The Gaseous State: Distribution of Molecular Velocities; Average Molecular Speed; Law of Partial Pressures; Mean Free Path—Macro- and Micro-states—The Quantum Theory: Energy Distribution Formulae; Quantization—The Atomic Theory: The Rutherford-Bohr Atom; Excitation Potentials; The Magneton; Elliptical Orbits; Spatial Quantization; Electron Spin; The Vector Model Atom; Atomic Structure; Pauli's Exclusion Principle

CHAPTER II

GASEOUS ELECTRICAL CONDUCTION 57

Methods of Current Measurement—Theory of Conduction through Gases: Motion of Ions—Paschen's Law—Breakdown in Uniform Fields: Non-uniform Fields—Principle of Similarity—The Corona Discharge—The Spark Discharge: Time-lag—Streamer Theory of Spark Discharge—The Glow Discharge—The Arc Discharge—The Mercury Arc—Probes and Probe Measurements: Electron Temperature

CHAPTER III

THE ELECTRON 99

Cathode Rays—Mass, Charge, and Velocity of the Electron—The Wilson Cloud Chamber—Determination of the Electronic Charge: Thomson's, Wilson's, and Millikan's Methods; Electron Dimensions—Electron Dynamics: Variation of Mass with Velocity; Electromagnetic Momentum and Mass—Electron Theory of Physical Constants: Specific Inductive Capacity and Refractive Index; Refraction and Dispersion; Paraelectricity; Magnetism; Diamagnetism, Paramagnetism and Ferromagnetism; Explanation of Remanence

CHAPTER IV

ELECTRONS IN METALS 149

Classical Theory of Conductivity: Thermal Conductivity—Electromagnetic Induction—Thermionics—The Fermi-Dirac Theory of the Electron Gas: Average Electron Energy; Velocity Distribution of Emitted Electrons—Emission Constants and their Determination: Cathode Temperature Determination; Emission Efficiency—Contact Potentials—Field Emission—Photo-electricity: Millikan's Experiments; Einstein's Theory; Light Quanta—Secondary Emission—Conduction in Metals, Insulators, and Semi-conductors

CHAPTER V

PAGE

X-RAYS 197

Methods of Production: Secondary Radiation—Origin of X-rays: Classical and Quantum Theories; Absorption—Quantum Theory of Scattering—Secondary Corpuscular Radiation—Characteristic X-rays—X-ray Diffraction: Crystal Structure; Grating Space; The X-ray Spectrometer—X-rays and Atomic Number—X-ray Tubes—Applications of X-rays: Radiography

CHAPTER VI

ELECTRON OPTICS 227

Refraction of Electrons: Refractive Index—The Single-diaphragm Aperture Lens—Three-diaphragm Aperture Lens—The Two-diaphragm Lens: Linear Magnification—Potential Distribution—Electron Trajectories—Magnetic Focusing: Short Solenoid Focusing; Focal Length; Magnification—Space Charge Effects—Gas Lenses

CHAPTER VII

LUMINESCENCE 259

Photo-luminescence: Stoke's Law; Excitation of Luminescence; Efficiency—Cathode-luminescence: Permanence of Luminescence—Luminescent Materials: Activation and Quenching—Electronic Applications of Luminescence: X-ray Screens; Oscillograph and Television Screens—The Electron Microscope—Fluorescent Lamps

CHAPTER VIII

HIGH-VACUUM DIODES 273

Space-charge Equations—Effect of Cathode Potential Drop—A.C. Directly-heated Cathodes—Modifying Factors of the Space-charge Equation—Vacuum Valve Design and Construction—Cathode Materials: Tungsten Cathodes; Thoriated Tungsten Cathodes—Oxide-coated Cathodes; Indirectly-heated Cathodes; Heat-shielded Cathodes—The Anode—Electrode Leads

CHAPTER IX

GAS-FILLED DIODES AND TRIODES 290

Conduction Conditions in Gas-filled Thermionic Diodes: Valve Volt Drop; Influence of Gas Pressure—Cathodes in Gas-filled Valves: Cathode Emission Efficiency—Anode Temperature—Thyratrons: Theory of Striking; Grid Current; Grid Action Before Striking; Grid Action After Striking; Ionization and De-ionization Times; Influence of Gas Pressure on Thyratron Characteristics

CHAPTER X

HIGH-VACUUM THERMIONIC TRIODES AND MULTI-ELECTRODE VALVES 312

Mutual Characteristics: Amplification Factor; Triode Characteristics—The Triode as a Circuit Element: Triode Input Impedance; Input Impedance at Radio Frequencies—Rectifying Properties of Triodes:

PAGE

Cumulative Grid Rectification—Multi-electrode Valves: Tetrode Characteristics; The Pentode; Variable-μ Valves; The Octode Frequency Changer

CHAPTER XI

VALVE AMPLIFIERS 338

Voltage Amplification Factor—Resistance-capacitance Coupled Amplifiers—Choke-capacitance Coupling—Tuned Anode Coupling—Transformer Coupled Amplifiers—High-frequency Transformer Coupling—Distortion in Amplifiers: Load Line with Reactive Load—Power Amplification: Graphical Determination of Power; Optimum Value of Grid Bias—Parallel-feed Impedance Matching—Anode Circuit Efficiency—Amplification with Pentodes—Push-pull Amplification: Composite Anode Characteristics—Class B Amplification: The Driver Stage—Class AB Amplification

CHAPTER XII

FEED-BACK. VALVE OSCILLATORS 379

Negative Feed-back: Influence on Harmonic Distortion—Negative Feed-back Circuits: Compound Feed-back; Positive Feed-back—Valve Oscillators: Tuned Grid and Tuned Anode Oscillators; Power in Oscillatory Circuit; Anode Tapping Point; Grid Excitation; Feed; Grid Bias; Oscillation Amplitude—Further Types of Oscillators: Negative Resistance Oscillators; The Magnetron Oscillator

CHAPTER XIII

ELECTRONIC RECTIFIERS 399

Asymmetrical Conductivity—Contact Rectifiers: The Electrolytic Rectifier; The Copper-oxide Rectifier; The Barrier Layer; Rectifier Resistance; Current and Temperature Effects; Electrical Properties of Copper-oxide Rectifiers; Creep; Ageing—The Selenium Rectifier—The Hot-cathode Rectifier: Cross-firing—High-vacuum Rectifiers—The Mercury-arc Rectifier: Sealing; Vacuum Maintenance—The Ignitron—Rectifier Losses—Rectifier Voltage and Current Relationships—Utility Factor: Overlap—Arc and Resistance Drops

CHAPTER XIV

RECTIFIER CIRCUITS 452

Single-phase Half-wave Circuits—Single-phase Full-wave Circuits—The Bridge Circuit—Three-phase Rectification: The three-phase Inter-star Circuit—Six-phase Rectification: The Delta Double-star Circuit; The Six-phase Fork Circuit; The Six-phase Bridge Circuit—Phase Equalizing: Interphase reactor—Twelve-phase Rectification: The Twelve-phase Fork Circuit; The Quadruple Three-phase Circuit

CHAPTER XV

GRID-CONTROLLED RECTIFIERS, INVERTERS, AND FREQUENCY CHANGERS 489

Soft-control Systems—Phase-shift Control: Phase-shifting Circuits—Bias-shift Control: Grid-controlled Polyphase Rectifiers; Ignition Angle; Circuit Interruption by Grid Control—Hard Control

Systems—The Mercury-Arc Inverter—Applications of the Mercury-arc Inverter: Regeneration; Static Frequency Changing; Reactive Power; The Envelope Frequency Convertor

CHAPTER XVI

PAGE

ELECTRONIC MEASURING INSTRUMENTS 519

Valve Voltmeters—Peak Voltmeters: The Slide-back Voltmeter; The Ryall Crest Voltmeter—Electronic Ammeter—The Electronic Wattmeter—Electronic Ohmmeter—Electronic Harmonic Analysis

CHAPTER XVII

CATHODE-RAY TUBES AND ASSOCIATED CIRCUITS . . . 542

Gas-focused Tubes: Origin Distortion; High-frequency Defocusing—Medium-voltage High-vacuum Tubes: Vacuum Tube Distortion; Trapezium Distortion; Symmetrical Deflexion—The High Voltage Tube: Beam Deflexion; Magnetic Deflexion; Effect of Electron Transit Time—Time Bases: The Thyratron Time Base; Hard-valve Time Bases—Synchronization—Circular and Elliptical Time Bases: Waveform Injection—Oscillograph Construction: Signal Amplification

CHAPTER XVIII

PHOTO-ELECTRIC CELLS 589

Photo-emissive Cells—Vacuum Photo-cells—Cathode Materials: Sensitized Cathodes; Cathode Films—Gas-filled Photo-cells: Spectral Response Curves—Photo-cell Applications—The Multiplier Photo-cell—Photo-voltaic Cells: Theory of Cell Operation; Effect of Temperature—Photo-conducting Cells: Cell Resistance; Time Lag

CHAPTER XIX

ELECTRIC DISCHARGE LAMPS 620

Temperature Radiation—Luminescent Radiation—Low-pressure Cold-cathode Discharge Lamps: The Rare Gases; The Cathode Fall; The Positive Column—Transformer Operation of Neon Lamps—Operating Characteristics of Neon Lamps: Inductive Control; Voltage and Current Waveforms—High-pressure Mercury-vapour Lamps—The Sodium-vapour Lamp—Fluorescent Lamps: Operating Circuits and Auxiliary Equipment—Mathematics of Discharge-lamp Circuits: Resistive, Capacitive, and Inductive Control; Power Factor Improvement

CHAPTER XX

MISCELLANEOUS ELECTRONIC DEVICES 665

The Cyclotron—The Betatron—The Electron Microscope: Specimen Mounting; Limitations of the Electron Microscope—The Cathode-ray Tuning Indicator—Voltage Stabilization with Discharge Tubes—Electronic Voltage Regulators—Electronic Control of Temperature

INDEXES 691

ELECTRONICS

CHAPTER I

ELECTRONICS AND THE FINE STRUCTURE OF MATTER—THE KINETIC, QUANTUM, AND ATOMIC THEORIES

Introduction

BROADLY speaking, two methods are employed (and are necessary) for the study of electrical phenomena. The first method deals with macroscopic quantities, experimental laws relating to these, and their mathematical consequences. Generally, the method is free from hypothesis, no assumptions being made concerning the structure of electrical charges or matter. This method of approach is termed the field theory, the subjects coming within its particular range being, principally, electrostatics and electrodynamics. A wide range of phenomena is, however, largely incapable of treatment by this method; e.g. electrolytic and gaseous conduction, the origin of radiation and spectra, thermionics, the photo-electric effect, electron optics, etc. In order to study these important phenomena assumptions must be made concerning the structure of matter and electrical charges, assumptions which, nevertheless, cannot be verified by direct methods. This second method, which is essential for the study of electronic matters, may be termed the atomistic method. In contrast to the field theory, which employs macroscopic concepts, the atomistic theory makes use of microscopic concepts, and hence it will be found that electronics is closely connected with the finer states of matter.

Electronics, as the term implies, is concerned with the science and applications of the electron. In particular it concerns the study of electron behaviour in magnetic and electric fields or combinations of both. However, electrons are seldom found unassociated with their parent atoms or ions and hence the latter must be included in any study of electronic matters. In all practical applications electrons and ions are invariably found in enormous numbers, and it is to be anticipated that in many instances their behaviour will be governed by laws appertaining to matter in a gaseous state. This is so, and hence it follows that some knowledge of the principal gas laws is indispensable to the study of electronics.

The Origin of Electrons

The electron is the natural unit of electricity possessing a charge of $1{\cdot}59 \times 10^{-19}$ coulomb, a mass of 9×10^{-28} gram, and, on the assumption of a spherical form, a radius of $1{\cdot}9 \times 10^{-13}$ cm. It is normally associated with the atom of which it forms an essential part. As matter is normally electrically neutral, it follows that the resultant negative charge of the electrons within an atom must be balanced by an equal positive charge. This necessity, of course, creates speculation concerning the structure of the atom, the consideration of which, however, will be left until later.

Atoms and Molecules

The atom may be defined as the smallest particle of matter capable of independent existence. This does not mean that an atom can in no way be divided, but rather that should this be effected then will the atom no longer possess those "properties" which are characteristic of the particular kind of substance to which it normally belongs. The atom is an indirectly observable object with certain properties ascertainable by experiment and others which exist mainly in imagination. It is practically certain, however, that it consists of a complicated configuration of electric charges, those of negative character being due to electrons, while the positive charge is associated with the relatively massive nucleus. Practically the whole mass of the atom resides in the latter, the electrons contributing a negligible amount to the total mass. Although the masses of different atoms differ considerably, their radii are similar and are of the order of 10^{-8} cm. In practice, atoms generally associate in groups of two or more, such groups being known as molecules. Where the atoms forming a molecule are different, then a substance composed of such molecules is termed "compound." When the atoms are similar the substance is "elementary." Thus, molecules and substances may be classified as either elements or compounds, a molecule being defined as the smallest particle of a compound capable of independent existence.

In the majority of applications of electrons the latter interact with matter in a gaseous state. Examples of this are the mercury-arc rectifier, the thyratron, gas-discharge lamps, photo-electric cells, "vacuum" valves, etc. In all these instances the electrons have a directed motion through the medium by virtue of the existence of an electric field, this motion, of course, constituting an electric current. In general the electrons tend to move along the lines of force of the field, but make frequent collisions with the gas molecules constituting the medium. It is evident that the latter may play an important role in electronic conduction, particularly as its influence can be profoundly modified by changes in temperature, density, etc. Hence before proceeding further it is necessary to consider some properties of gaseous media.

The Gaseous State

The classical concept of a perfect gas is one of myriads of infinitesimally small particles (molecules or atoms) in a state of incessant motion. Frequent collisions or encounters occur between the particles, which are regarded as behaving as perfectly elastic solids with a coefficient of restitution of unity. Between encounters the motion is rectilinear and it is assumed that the time of an encounter is negligible compared with the time between encounters. The average distance travelled by a particle between two consecutive collisions is extremely small, being of the order of 10^{-5} cm. in gases at normal temperature and pressure. Under the same circumstances the collision frequency is about 10^{10} per sec. Statements such as these enable a picture to be formed of the environment in which electrons and ions find themselves when participating in a gaseous electrical discharge such as a glow or arc, etc. Furthermore, in the case of an electrical breakdown in a gaseous medium it will be found that the average distance between collisions (or the mean free path) is the principal determining factor of the breakdown potential.

For gases under normal conditions the volume of the molecules is negligible compared with the volume of any circumscribing space. This is evidenced by a comparison of the volumes occupied by a substance when in the gaseous and solid (or liquid) states. For example, a given quantity of water vapour at 100° C. and 760 mm. pressure occupies only 1/1600 of its volume as vapour when condensed to water at 100° C. From such considerations it follows that the dimensions of a molecule are negligible compared with the average distance traversed between collisions.

In employing the term "collision" to denote encounters between molecules, this term is not necessarily to be interpreted in its macroscopic sense. If powerful repulsive forces come into play (as is likely) when the centre of a molecule penetrates the sphere of influence of another molecule, this is sufficient to explain the sudden and continual changes in velocity which occur within a gas. Such forces are probably due to the interpenetration of the electron shells of the atoms.

BOYLE'S LAW

It is found that gases far removed from their liquefaction point approximately obey the law

$$pv = RT \quad . \quad . \quad . \quad . \quad (1\text{–}1)$$

Where p, v, and T are respectively the gas pressure, volume; and temperature, and R is a constant which, when referred to one gram-molecule, is termed the universal gas constant. The law is most nearly obeyed by the so-called "permanent" gases such as air, hydrogen, nitrogen, etc. A useful conception is, therefore, an ideal gas which obeys (1–1) exactly. For gases near their liquefaction point Boyle's Law, i.e. (1–1), no longer holds, and their behaviour is more nearly represented by the equation of Van der Waals, viz.

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT \quad . \quad . \quad . \quad (1\text{–}2)$$

where a and b are constants. Thus (1–1) and (1–2) may be respectively regarded as the equations of state of ideal and real gases.

With certain foregoing assumptions concerning molecular behaviour, and others shortly to be stated, relations may now be deduced between the macroscopic quantities p, v, T, and the mass, velocity, collision frequency, etc., of the molecules.

For the purposes of mathematical treatment the molecules may be regarded as elastic spheres behaving according to the laws of ordinary mechanics. Considering a volume of gas in thermal equilibrium, the number of contained molecules is so large that any sample taken at random will appear to be identical with any other sample. This means that, on the average, the molecular density is uniform throughout the gas. Due to constant encounters between the molecules the gas attains a state of statistical equilibrium or molecular chaos. This means that in any number of samples of the gas there will be the same distribution of molecular speeds, the same average speed, and the same mean free path between collisions.

It is evident that even if the molecules in a given volume actually possessed the same velocity at any instant, this uniformity would be rapidly destroyed by molecular collisions. Thus in a short time it would be found that the molecules were moving in all directions with every possible velocity. It is often convenient to represent graphically the velocity c of a molecule by a point in space whose co-ordinates are equal to the components of its velocity in three mutually perpendicular directions. Such a point is termed a velocity-point, and the space in which it is drawn the velocity-space.

This is represented in Fig. 1–1 where c may be called the velocity-point corresponding to any particular molecule. If the velocity-points of all the molecules in unit volume of gas are marked on the diagram, then the density of the points will vary with the distance from the origin, but since all directions are equally probable the density of the velocity-points in the velocity-space will be symmetrical about the origin. This means that the density is a function of c only.

Distribution of Molecular Velocities

Referring to Fig. 1–1, let u, v, w, be the velocity components along the x, y, z axes of a molecule moving with velocity c. Then

$$u^2 + v^2 + w^2 = c^2 \quad . \quad . \quad . \quad (1–3)$$

Let n_u be the number of molecules per cubic centimetre whose maximum velocity difference is unity at velocity u. Then if n is the number of molecules per cubic centimetre, $n_u/n = f(u)$ is the probability of a molecule having a velocity of $(u \pm \frac{1}{2})$ cm. per sec. The probability of a molecule having a velocity lying between u and $u + du$ is $f(u)du$, and the number of molecules per cubic centimetre having velocities between these limits is $nf(u)du$. In a similar manner the probability of a molecule having velocities between v and $v + dv$ and w and $w + dw$ is $f(v)dv$ and $f(w)dw$ respectively. Now as these three velocities are mutually perpendicular they do not affect each other and hence are independent. Thus the probability of a molecule experiencing the three simultaneously is given by the product of the probabilities; i.e. by

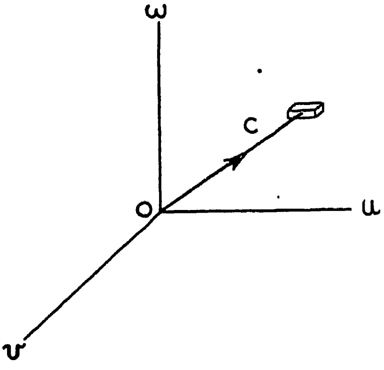


Fig. 1–1

$$f(u)f(v)f(w)dudvdw \quad . \quad . \quad . \quad (1–4)$$

But $dudvdw$ is a small volume in the velocity-space at a distance c from the origin. Hence (1–4) is the probability of a molecule finding itself within this space, and thus the number of molecules contained in the volume $dudvdw$ is

$$nf(u)f(v)f(w)dudvdw \quad . \quad . \quad . \quad (1–5)$$

However, on referring to the velocity diagram it is apparent that the number of velocity-points (or molecules) lying within a parallelopiped $dudvdw$ depends only on the volume of this and its distance, c, from the origin. We may then write

$$nf(u)f(v)f(w)dudvdw = nF(c)dudvdw$$

or

$$f(u)f(v)f(w) = F(c) = \phi(c^2) = \phi(u^2 + v^2 + w^2) \qquad (1\text{–}6)$$

where F is a function of c and ϕ a different function of c^2. Now when the gas is in a steady state the number of velocity-points within the volume $dudvdw$ is independent of time; i.e. as many molecules enter the volume as leave in a given period. Hence the rate of change of (1–5) with respect to time is zero. Taking n and $dudvdw$ as constant and differentiating, we have

$$f(v)f(w)f'(u)\frac{du}{dt} + f(u)f(w)f'(v)\frac{dv}{dt} + f(u)f(v)f'(w)\frac{dw}{dt} = 0$$

or

$$\frac{f'(u)}{f(u)}du + \frac{f'(v)}{f(v)}dv + \frac{f'(w)}{f(w)}dw = 0 \quad . \quad . \quad (1\text{–}7)$$

Now for those molecules whose speeds lie between c and $c + dc$ (1–3) is constant. Hence differentiating c with respect to time

$$udu + vdv + wdw = 0$$

Multiplying this by an undetermined factor $2/a^2$ and adding to (1–7) we have

$$\left(\frac{f'(u)}{f(u)} + \frac{2u}{a^2}\right)du + \left(\frac{f'(v)}{f(v)} + \frac{2v}{a^2}\right)dv + \left(\frac{f'(w)}{f(w)} + \frac{2w}{a^2}\right)dw = 0 \quad (1\text{–}8)$$

As each of the terms in brackets in (1–8) is independent of the others, we may put

$$\left(\frac{f'(u)}{f(u)} + \frac{2u}{a^2}\right)du = 0$$

$$\left(\frac{f'(v)}{f(v)} + \frac{2v}{a^2}\right)dv = 0$$

$$\left(\frac{f'(w)}{f(w)} + \frac{2w}{a^2}\right)dw = 0$$

which leads to

$$f(u) = K\varepsilon^{-\frac{u^2}{a^2}} \quad . \quad . \quad . \quad . \quad (1\text{–}9)$$

$$f(v) = K\varepsilon^{-\frac{v^2}{a^2}} \quad . \quad . \quad . \quad . \quad (1\text{–}10)$$

$$f(w) = K\varepsilon^{-\frac{w^2}{a^2}} \quad . \quad . \quad . \quad . \quad (1\text{–}11)$$

$$\phi(c^2) = K^3\varepsilon^{-\frac{u^2 + v^2 + w^2}{a^2}} \quad . \quad . \quad . \quad (1\text{–}12)$$

where K is a constant. (1–12) is known as the Maxwell Law of Distribution of Velocities.

As $f(u) = K\varepsilon^{-\frac{u^2}{a^2}}$ the number of molecules per cubic centimetre having velocities between u and $u + du$ is

$$nK\varepsilon^{-\frac{u^2}{a^2}}du \quad . \quad . \quad . \quad . \quad (1\text{–}13)$$

and hence the integration of this expression between the limits $+\infty$ and $-\infty$ will give the number of molecules per cubic centimetre. Thus

$$\int_{-\infty}^{+\infty} nK\varepsilon^{-\frac{u^2}{a^2}}du = n$$

or

$$K\int_{-\infty}^{+\infty} \varepsilon^{-\frac{u^2}{a^2}}du = 1 \quad . \quad . \quad . \quad (1\text{–}14)$$

from which we shall seek to determine K.

Similar expressions to that of (1–14), of course, exist for the v and w velocity components. Taking the expression in v and multiplying by (1–14), we obtain

$$K^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \varepsilon^{-\frac{u^2+v^2}{a^2}}dudv = 1 \quad . \quad . \quad (1\text{–}15)$$

Transforming into polar co-ordinates, r and θ,

$$u^2 + v^2 = r^2$$

and

$$dudv = rdrd\theta$$

Substituting in (1–15)

$$K^2\int_{0}^{\infty}\int_{0}^{2\pi} \varepsilon^{-\frac{r^2}{a^2}}rdrd\theta = 1$$

and performing the first integration, there results

$$2\pi K^2\int_{0}^{\infty} \varepsilon^{-\frac{r^2}{a^2}}rdr = 1 \quad . \quad . \quad . \quad (1\text{–}16)$$

The differential of $\varepsilon^{-\frac{r^2}{a^2}}$ being $-\frac{2}{a^2}\varepsilon^{-\frac{r^2}{a^2}}rdr$ the integration of (1–16) gives

$$-a^2\pi K^2\left[\varepsilon^{-\frac{r^2}{a^2}}\right]_0^\infty$$

$$= a^2\pi K^2 = 1$$

from which

$$K = \frac{1}{a\sqrt{\pi}} \quad . \quad . \quad . \quad . \quad (1\text{–}17)$$

Hence (1–9) becomes

$$\frac{1}{a\sqrt{\pi}}\varepsilon^{-\frac{u^2}{a^2}}$$

and thus the number of molecules having velocities between the limits u and $u + du$ is

$$\frac{n}{a\sqrt{\pi}}\varepsilon^{-\frac{u^2}{a^2}}du \quad . \quad . \quad . \quad . \quad (1\text{–}18)$$

Referring to (1–12) the number of molecules whose velocity-points lie within the volume $dudvdw$ is

$$\frac{n}{a^3\pi^{3/2}}\varepsilon^{-\frac{(u^2+v^2+w^2)}{a^2}}dudvdw = \frac{n}{a^3\pi^{3/2}}\varepsilon^{-\frac{c^2}{a^2}}dudvdw$$

Now $dudvdw$ is an elementary volume in the velocity-space, and it follows that the number of velocity points, or molecules, within a thin spherical shell of radius c and thickness dc is

$$\frac{n}{a^3\pi^{3/2}}\varepsilon^{-\frac{c^2}{a^2}}4\pi c^2dc = \frac{4n}{a^3\sqrt{\pi}}\varepsilon^{-\frac{c^2}{a^2}}c^2dc \quad . \quad . \quad (1\text{–}19)$$

(1–19), of course, gives the number of molecules whose speeds lie between c and $c + dc$.

The curve whose equation is

$$y = \frac{4}{a^3\sqrt{\pi}}\varepsilon^{-\frac{c^2}{a^2}}c^2 \quad . \quad . \quad . \quad (1\text{–}20)$$

is shown by Fig. 1–2, and is known as a probability curve. The ordinate at any point gives the value of $\phi(c^2)$ corresponding to a particular value of c and, if two vertical lines AB and CD are erected at distances c and $c + dc$ from the origin, the area enclosed between them, the base dc, and the ordinate $\phi(c^2)$ will represent the number of molecules having speeds between c and $c + dc$. The maximum of the curve corresponds to a speed known as the "most probable speed," this term being employed because, for a given speed difference, dc, a larger percentage of molecules have this speed than any other. In order to find the most probable speed (1·20) is differentiated and equated to zero. This gives

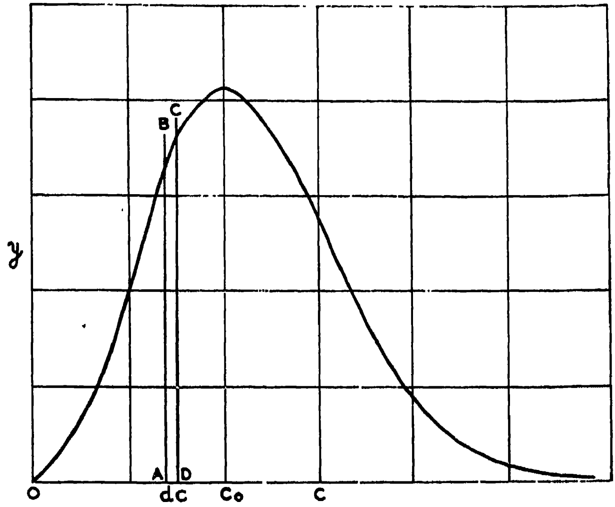


FIG. 1–2

$$C_0 = a$$

i.e. a is the most probable speed.

AVERAGE SPEED OF THE MOLECULES

In order to find the average speed of the molecules we must sum the speeds of all molecules and then divide by the total number of molecules participating in those speeds. Hence we multiply (1–19) by c, divide by n, and integrate as below, where $\overline{C}$ is the mean speed.

$$\overline{C} = \frac{1}{n}\int_0^\infty c \times \frac{4n}{a^3\sqrt{\pi}}\varepsilon^{-\frac{c^2}{a^2}}c^2dc = \frac{4}{a^3\sqrt{\pi}}\int_0^\infty \varepsilon^{-\frac{c^2}{a^2}}c^3dc \quad . \quad (1\cdot21)$$

Integrating the member to the right of the sign of integration by parts

$$\int_0^\infty \varepsilon^{-\frac{c^2}{a^2}} c^3 dc = \int_0^\infty c^2 \,.\, c\varepsilon^{-\frac{c^2}{a^2}} dc$$

$$= -\frac{a^2}{2}\left[c\varepsilon^{-\frac{c^2}{a^2}}\right]_0^\infty + \frac{2a^2}{2}\int_0^\infty c\varepsilon^{-\frac{c^2}{a^2}} dc$$

$$= \frac{a^4}{2}$$

which gives for (1–21)

$$\overline{C} = \frac{2a}{\sqrt{\pi}}$$

A further speed of considerable importance is the average value of squares of the speeds. This is expressed by

$$C^2 = \frac{1}{n}\int_0^\infty c^2 \times \frac{4n}{a^3\sqrt{\pi}}\varepsilon^{-\frac{c^2}{a^2}} c^2 dc = \frac{4}{a^3\sqrt{\pi}}\int_0^\infty \varepsilon^{-\frac{c^2}{a^2}} c^4 dc$$

Integrating by parts

$$\int_0^\infty \varepsilon^{-\frac{c^2}{a^2}} c^4 dc = \int_0^\infty c^3 \,.\, c\varepsilon^{-\frac{c^2}{a^2}} dc$$

$$= -\frac{a^2}{2}\left[c^3\varepsilon^{-\frac{c^2}{a^2}}\right]_0^\infty + \frac{3a^2}{2}\int_0^\infty c^2\varepsilon^{-\frac{c^2}{a^2}} dc$$

$$= 3 \times \frac{a^2}{2} \times \frac{a^2}{2}\left[c\varepsilon^{-\frac{c^2}{a^2}}\right]_0^\infty + \frac{3}{2} \times \frac{a^4}{2}\int_0^\infty \varepsilon^{-\frac{c^2}{a^2}} dc$$

$$= \frac{3a^4}{4}\int_0^\infty \varepsilon^{-\frac{c^2}{a^2}} dc \quad . \quad . \quad . \quad . \quad . \quad (1\text{–}22)$$

This integral may be evaluated by a similar device as that employed on page 7, but as the limits here are 0 and ∞, the corresponding limits in polar co-ordinates must be such as to cover one quadrant only. Hence (1–22) becomes

$$\frac{3a^4}{4} \times \frac{a}{2}\sqrt{\pi}$$

which leads to

$$C^2 = \frac{4}{a^3\sqrt{\pi}} \times \frac{3a^4}{4} \times \frac{a}{2}\sqrt{\pi} = \frac{3}{2}a^2 \quad . \quad . \quad (1\text{–}23)$$

Summarizing we have

$$C_0 = a \quad . \quad . \quad . \quad . \quad (1\text{–}24)$$

$$\bar{C} = 2a/\sqrt{\pi}\ . \quad . \quad . \quad . \quad (1\text{–}25)$$

$$C = \sqrt{3/2}a\ . \quad . \quad . \quad . \quad (1\text{–}26)$$

$$\bar{C}/C = \sqrt{8/3\pi}\ . \quad . \quad . \quad . \quad (1\text{–}27)$$

From the foregoing it is evident that the determination of molecular velocities and speeds depends on the as yet unidentified parameter a. This will now be determined in terms of macroscopically observable quantities pressure, volume, temperature, etc.

Consider a cubical container with a length of side equal to 1 cm. The walls of the container are assumed to be perfectly plane with a coefficient of restitution of unity, so that a molecule striking a wall has its velocity component perpendicular to that wall exactly reversed. If the mass of the molecule is m and u is the velocity component before impact, it will be $-u$ after impact. Hence the change of momentum is $2mu$. In the absence of molecular collisions, a molecule would travel to and fro u times per second within the container, striking against a wall $u/2$ times per second. Hence in these circumstances the change of momentum per second per molecule is

$$2mu \times u/2 = mu^2$$

and the change of momentum per second, p, due to n molecules, is

$$p = m(u_1^2 + u_2^2 + u_3^2 + \ldots u_n^2) \quad . \quad . \quad . \quad (1\text{–}28)$$

Similarly for the walls perpendicular to the v and w velocity components

$$p = m(v_1^2 + v_2^2 + v_3^2 + \ldots v_n^2) \quad . \quad . \quad (1\text{–}29)$$

$$p = m(w_1^2 + w_2^2 + w_3^2 + \ldots w_n^2)\ . \quad . \quad (1\text{–}30)$$

Adding (1–28), (1–29), and (1–30)

$$3p = m(c_1^2 + c_2^2 + c_3^2 + \ldots c_n^2) \quad . \quad . \quad (1\text{–}31)$$

Due to collisions c_1, c_2, etc., are, of course, different, the various equations above taking this into account. If C^2 is the mean of the squares of the speeds in (1–31), then this becomes

$$3p = mnC^2$$

and
$$C^2 = 3p/mn$$

But mn is the density of the gas, ρ, and hence

$$C^2 = 3p/\rho \quad . \quad . \quad . \quad . \quad (1\text{–}32)$$

$$C = \sqrt{3p/\rho} \quad . \quad . \quad . \quad . \quad (1\text{–}33)$$

From (1–26) we also have

$$C = \sqrt{3/2}a$$

and thus
$$a = \sqrt{\frac{2p}{\rho}}$$

The result given by (1–32) may also be obtained with the aid of the expression for the distribution of the velocities among the molecules. Thus, the change of momentum per second due to molecules having velocity components between u and $u + du$ is

$$\frac{n}{a\sqrt{\pi}} \varepsilon^{-\frac{u^2}{a^2}} mu^2du$$

and thus the total change in momentum per second due to all molecules is

$$p = \frac{nm}{a\sqrt{\pi}} \int_0^\infty \varepsilon^{-\frac{u^2}{a^2}} u^2du$$

$$= \frac{nm}{a\sqrt{\pi}} \cdot \frac{a^3\sqrt{\pi}}{2} = \frac{1}{2} nma^2$$

or
$$p = \frac{1}{3} nmC^2$$

and
$$C^2 = 3p/mn, \text{ as before}$$

Now if V_0 represents the molecular volume of a gas and M its molecular weight $\rho = M/V_0$, and from (1–32)

$$C^2 = \frac{3p}{M/V_0}$$

and
$$pV_0 = \frac{MC^2}{3} \quad \text{(Boyle's Law)}$$

From (1–1) we have

$$\frac{MC^2}{3} = RT$$

and calling N the ratio M/m, i.e. the number of molecules in a gram-molecule,

$$\frac{mC^2}{3} = \frac{RT}{N}$$

and

$$\frac{1}{2}mC^2 = \frac{3RT}{2N} = \frac{3}{2}kT = \alpha T \quad . \quad . \quad . \quad (1\text{–}34)$$

where $\alpha = 3R/2N$.

From this result we must consider the absolute temperature of a gas as proportional to the mean kinetic energy of translation of its molecules. Thus to every uniform temperature there is a corresponding uniform distribution of kinetic energy, molecular collisions furnishing the means which lead to an equalization of temperature.

LAW OF PARTIAL PRESSURES

The law expressed by (1–31) relates to a gas whose molecules are all identical. Assume now that a space contains a mixture of gases, the number of molecules of each per cubic centimetre being denoted by n_1, n_2, n_3, etc., and their respective masses by m_1, m_2, m_3, etc. From (1–31) we have

$$p = \tfrac{1}{3}\sum_1^n mc^2$$

and hence for the mixture

$$\begin{aligned} p &= \tfrac{1}{3}\sum_1^{n_1} m_1c^2 + \tfrac{1}{3}\sum_1^{n_2} m_2c^2 + \tfrac{1}{3}\sum_1^{n_3} m_3c^2 + \text{etc.} \\ &= p_1 + p_2 + p_3 + \text{etc.} \end{aligned}$$

Or the total pressure of a mixture of gases is the sum of the separate pressures which each would exert if acting alone. This is usually referred to as Dalton's Law.

The result of collisions between molecules of different kinds will now be considered with regard to the distribution of kinetic energy among the various different molecules.

Let a collision occur between two molecules whose masses are respectively m_1 and m_2. Further, let the line joining their centres on collision be parallel to the axis of the u-component of the velocity.

Then if u and u' are the respective velocities of m_1 and m_2 before collision and similarly U and U' are the velocities after collision

$$m_1u + m_2u' = m_1U + m_2U' \quad . \quad . \quad (1\text{–}35)$$

$$m_1u^2 + m_2u'^2 = m_1U^2 + m_2U'^2 \quad . \quad . \quad (1\text{–}36)$$

Solving for U and U' we obtain

$$U = \frac{2m_2u' + (m_1 - m_2)u}{m_1 + m_2} \text{ or } u$$

$$U' = \frac{2m_1u + (m_2 - m_1)u'}{m_1 + m_2} \text{ or } u'$$

Excluding the second pair of values (that is, where there is no change of velocity), the difference in the kinetic energies of the molecules after collision is

$$\tfrac{1}{2}m_1U^2 - \tfrac{1}{2}m_2U'^2$$

$$= \left(\frac{8m_1m_2}{(m_1 + m_2)^2} - 1\right)(\tfrac{1}{2}m_2u'^2 - \tfrac{1}{2}m_1u^2) + \frac{4m_1m_2(m_1 - m_2)uu'}{(m_1 + m_2)^2} \quad (1\text{–}37)$$

Now (1–37) does not indicate whether the difference in the kinetic energies of a colliding pair of molecules is increased or decreased by collision. However, it is the average effect taken over all the molecules which is of interest and in this case definite conclusions may be drawn.

Considering the second term of the right-hand member of (1–37), it will be noted that this differs for different pairs of molecules, solely by virtue of differences in uu'. Due to the symmetry of the velocity diagram, it is apparent that for any given magnitudes of u and u' there will be just as many cases where u and u' have the same signs as opposite signs. Hence in summing (1–37) for all the molecules the second term of (1–37) will disappear. The second factor of the first term is the difference of the energies of the molecules before collision, and hence

$$\frac{E_2}{E_1} = \frac{8m_1m_2}{(m_1 + m_2)^2} - 1 \quad . \quad . \quad . \quad (1\text{–}38)$$

where E_1 and E_2 are respectively the energy differences before and after collision. It is evident that the average difference of the energies after collision is greater or less than the average difference before, according to whether the absolute value of the right-hand member of (1–38) is greater or less than unity. Assuming that

$$\frac{8m_1m_2}{(m_1+m_2)^2}-1<1$$

then

$$\frac{4m_1m_2}{(m_1+m_2)^2}<1$$

$$4m_1m_2<(m_1^2+m_2^2+2m_1m_2)$$

$$0<(m_1^2+m_2^2-2m_1m_2)$$

$$0<(m_1-m_2)^2 \quad . \quad . \quad . \quad . \quad (1\text{–}39)$$

The result of assuming the right-hand member of (1–38) is greater than unity is found by reversing the sign of inequality in (1–39), which leads to an impossible conclusion.

We thus see that the difference between the average kinetic energies of the two sets of molecules tends to decrease with the number of collisions, and from this it may be concluded that when a mixture of gases is in equilibrium the average kinetic energies of each of the kinds of molecules will be the same.

That is $$\tfrac{1}{2}m_1C_1^2=\tfrac{1}{2}m_2C_2^2=\tfrac{1}{2}m_3C_3^2= \ . \ . \ . \ . \quad . \quad (1\text{–}40)$$

Under these conditions there is an equality of temperature throughout the mixture and, as we see from (1–34) that the temperature of a gas is proportional to the average kinetic energy of its molecules, it may be inferred that if the gases were separated, the temperature of each remaining constant, their average kinetic energies would be the same as when mixed. Hence (1–40) is true whether the gases are mixed or separate, provided the temperatures of all are the same. Thus, if different gases are at the same temperature and pressure,

$$p=\tfrac{1}{3}n_1m_1C_1^2=\tfrac{1}{3}n_2m_2C_2^2=\tfrac{1}{3}n_3m_3C_3^2= \ . \ . \ .$$

$$\tfrac{1}{2}m_1C_1^2=\tfrac{1}{2}m_2C^2=\tfrac{1}{2}m_3C_3^2= \ . \ . \ .$$

from which

$$n_1=n_2=n_3= \ . \ . \ .$$

or equal volumes of different gases under the same conditions of temperature and pressure contain the same number of molecules. This result is known as Avogadro's Law, and as we now see that N in (1–34) is the same for all gases so also must be α. Thus α is a universal constant.

THE MEAN FREE PATH

The average distance traversed by a molecule between two consecutive collisions is termed the *mean free path* of the molecules. A collision may be said to occur when the centres of the molecules approach to a distance of σ from each other, σ being the molecular diameter. Considering any particular molecule, let $\overline{C}_r$ be its mean velocity relative to all the other molecules. Then a collision occurs when any of the centres of the latter approach the surface of a sphere of radius σ surrounding the molecule under consideration. This sphere is termed the *sphere of molecular action.* The number of molecules whose centres will come within the sphere of molecular action of a moving molecule during a second is $\pi\sigma^2\overline{C}_r n$ as $\pi\sigma^2\overline{C}_r$ is the effective volume swept out by the molecule per second. Hence the collision frequency is

$$P = \pi\sigma^2 C_r n \qquad . \quad . \quad . \quad (1\text{–}41)$$

Now the actual distance travelled per second by a molecule is $\overline{C}$, and thus the average distance between collisions is

$$l = \frac{\overline{C}}{\pi\sigma^2\overline{C}_r n} = \frac{1}{\pi\sigma^2 n}\cdot\frac{\overline{C}}{\overline{C}_r} \qquad . \quad . \quad . \quad (1\text{–}42)$$

where l is the mean free path.

Now $\overline{C}$ is already known and to determine l we must therefore determine $\overline{C}_r$. To effect this, consider two sets of molecules whose velocities may be represented by c and c_1, and their relative velocity by c_r. Then the components may be represented as below

$$c^2 = u^2 + v^2 + w^2$$

$$c_1{}^2 = (u + \xi)^2 + (v + \eta)^2 + (w + \zeta)^2$$

$$c_r{}^2 = \xi^2 + \eta^2 + \zeta^2$$

Let the number of molecules of each kind per unit volume be given by n and n_1 and the most probable speeds by a and β. Then the respective number of each kind having x velocity components between u and $u + du$, and $u + \xi$, and $u + \xi + d\xi$ is

$$\frac{n}{a\sqrt{\pi}}\varepsilon^{-\frac{u^2}{a^2}}du \qquad . \quad . \quad . \quad . \quad (1\text{–}43)$$

and

$$\frac{n_1}{\beta\sqrt{\pi}}\varepsilon^{-\frac{(u+\xi)^2}{\beta^2}}d\xi \qquad . \quad . \quad . \quad (1\text{–}44)$$

Now the number of pairs of molecules, one of each kind, is given by the product of (1–43) and (1–44), viz.

$$\frac{nn_1}{\alpha\beta\pi}\varepsilon^{-\left[\frac{u^2}{\alpha^2}+\frac{(u+\xi)^2}{\beta^2}\right]}du\,d\xi$$

and the total number of such pairs may be found by integrating this expression for all values of u from $-\infty$ to $+\infty$, giving

$$\frac{nn_1 d\xi}{\alpha\beta\pi}\int_{-\infty}^{+\infty}\varepsilon^{-\left[\frac{u^2}{\alpha^2}+\frac{(u+\xi)^2}{\beta^2}\right]}du$$

or

$$I(\xi)=\varepsilon^{-\frac{\xi^2}{\beta^2}}\int_{-\infty}^{\infty}\varepsilon^{-\left[u^2\left(\frac{1}{\alpha^2}+\frac{1}{\beta^2}\right)+\frac{2u\xi}{\beta^2}\right]}du$$

Putting

$$\frac{\sqrt{\alpha^2+\beta^2}}{\alpha\beta}u=x$$

$$\frac{2\alpha\xi}{\beta\sqrt{\alpha^2+\beta^2}}=2b$$

then

$$I(\xi)=\varepsilon^{\left(b^2-\frac{\xi^2}{\beta^2}\right)}\frac{\alpha\beta}{\sqrt{\alpha^2+\beta^2}}\int_{-\infty}^{\infty}\varepsilon^{-(b+x)^2}d(b+x)$$

$$=\varepsilon^{\left(b^2-\frac{\xi^2}{\beta^2}\right)}\frac{\alpha\beta}{\sqrt{\alpha^2+\beta^2}}\sqrt{\pi}$$

Hence we have

$$\frac{nn_1\varepsilon^{\left(b^2-\frac{\xi^2}{\beta^2}\right)}}{\sqrt{\pi}\sqrt{\alpha^2+\beta^2}}d\xi=\frac{nn_1}{\sqrt{\pi}\sqrt{\alpha^2+\beta^2}}\varepsilon^{-\frac{\xi^2}{\alpha^2+\beta^2}}d\xi$$

Now this expression gives the number of pairs of molecules having relative speeds, the x-components of which lie between ξ and $\xi+d\xi$, the total number of possible pairs being nn_1. The expression is of the same form as (1–18), if we put nn_1 for n, $\alpha^2+\beta^2$ for α^2, and ξ for u. Thus, the distribution of relative velocities is governed by the same law as the velocities themselves. We have already seen that the mean speed of a molecule is given by

$$\overline{C}=2\alpha/\sqrt{\pi}$$

and hence it follows that the mean relative speed

$$\overline{C}_r=\frac{2\sqrt{\alpha^2+\beta^2}}{\sqrt{\pi}}$$

If the two sets of molecules form part of a common system, then $\beta = a$ and

$$\overline{C}_r = \frac{2a\sqrt{2}}{\sqrt{\pi}} = \overline{C}\sqrt{2}$$

from which
$$\frac{\overline{C}_r}{\overline{C}} = \sqrt{2}$$

and
$$l = \frac{1}{\sqrt{2}\pi\sigma^2 n} \quad . \quad . \quad . \quad . \quad (1\text{–}45)$$

$$P = \sqrt{2}\pi\sigma^2 n\overline{C} \quad . \quad . \quad . \quad . \quad (1\text{–}46)$$

The Distribution of Free Paths

The expression given by (1–45) is, of course, only the average distance travelled by a molecule between collisions. Actually the free paths follow a distribution law which will now be determined. The probability that a molecule will travel a distance x without collision may be written $f(x)$, and the probability that it will travel the distance $x + dx$ without collision $f(x + dx)$. Hence

$$f(x + dx) = f(x) + f'(x)dx$$

But the probability of the molecule covering both x and dx without collision is the product of the separate probabilities that both these events shall occur. Thus

$$f(x) + f'(x)dx = f(x)f(dx) \quad . \quad . \quad (1\text{–}47)$$

In order to determine $f(dx)$, consider a cylinder of unit cross-sectional area and length dx, the axis being parallel to the direction of motion of the molecule. Then, on the assumption that the molecules within the cylinder are at rest, the probability of a collision within the distance dx is the ratio of the sum of the projected areas of the spheres of influence of all the molecules to the cylindrical front surface of unit area. But we have already seen that due to the molecules being in motion the number of collisions in a given time is $\sqrt{2}$ times as great as when they are considered at rest. Hence the probability of collision is $\sqrt{2}\pi\sigma^2 ndx$ and the probability of covering the distance dx without collision is

$$f(dx) = 1 - \sqrt{2}\pi\sigma^2 ndx$$

From (1–47)

$$f(x) + f'(x)dx = f(x)(1 - \sqrt{2}\pi\sigma^2 n dx)$$

Integrating
$$\int \frac{f'(x)}{f(x)} dx = -\sqrt{2}\pi\sigma^2 n \int dx$$

and
$$f(x) = K\varepsilon^{-\sqrt{2}\pi\sigma^2 nx} = K\varepsilon^{-\frac{x}{l}}$$

as, from (1–45), $\sqrt{2}\pi\sigma^2 n = 1/l$. To determine K we note that the probability is unity for $x = 0$. Hence

$$f(x) = \varepsilon^{-\frac{x}{l}} \quad . \quad . \quad . \quad . \quad (1\text{–}48)$$

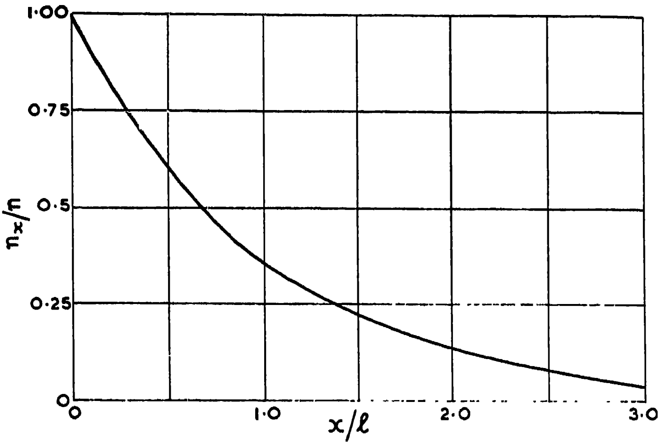


FIG. 1–3

Thus we may regard the mean free path as that path over which the probability of passage without collision is $\varepsilon^{-1} = 0{\cdot}366$. The expression given by (1–48) may also be written

$$\frac{n_x}{n} = \varepsilon^{-\frac{x}{l}} \quad . \quad . \quad . \quad . \quad (1\text{–}49)$$

where n_x is the number of molecules per unit volume which traverse a path x without collision. From (1–49) we note that only 36·6 per cent of the molecules travel a distance equal to the mean free path without collision. (1–49) is sometimes termed the survival equation, as it gives the number of molecules which have not suffered a collision in travelling a distance x. This equation is plotted in Fig. 1–3, from which we see that paths shorter than l are much more probable than those longer.

From (1–49) we may find the number of molecules, n_{dx}, which having survived collisions over a distance x, terminate their free paths within a further distance dx.

Thus
$$n_{dx} = dn = -\frac{n}{l}\varepsilon^{-\frac{x}{l}}\,dx \quad . \quad . \quad . \quad (1\text{–}50)$$

the negative sign indicating that n_{dx} decreases with an increase in x. The probability of survival is

$$\frac{n\varepsilon^{-\frac{x}{l}} - \frac{n}{l}\varepsilon^{-\frac{x}{l}}\,dx}{n\varepsilon^{-\frac{x}{l}}}$$
$$= 1 - \frac{dx}{l}$$

which is independent of x.

The expression given by (1–50) gives the number of molecules, dn, whose mean free paths lie between x and $x + dx$ and, because of this, it may be termed the distribution function of mean free paths. From (1–50) the sum of the lengths of all the paths is

$$\int_0^\infty \frac{xn}{l}\varepsilon^{-\frac{x}{l}}\,dx$$

and hence the mean free path should be the average of this, i.e.

$$\frac{1}{n}\int_0^\infty \frac{xn}{l}\varepsilon^{-\frac{x}{l}}\,dx$$
$$= [-x\varepsilon^{-\frac{x}{l}} + \int \varepsilon^{-\frac{x}{l}}]_0^\infty$$
$$= [-l\varepsilon^{-\frac{x}{l}}]_0^\infty = l$$

It will have been noted that the derivation of the velocity distribution function is based on the assumption that the components of the molecular velocity are independent. Because of this, and for other reasons, we shall now derive the distribution function by a more general method which does not depend on this assumption.

Macro- and Micro-states

In a similar manner to that expressed by (1–5) we may write
$$dn = F(u, v, w, t)\,du\,dv\,dw = F\,d\tau \quad . \quad . \quad (1\text{–}51)$$

where $d\tau$ represents a volume-element in the velocity-space. t represents time and is here included to emphasize that the law of distribution of velocities may vary with time. Comparing (1–51) with (1–5), we see that

$$F = nf(u)f(v)f(w)$$

The velocity-space is assumed to be split up into a number of identical volume-elements $d\tau$, the volume of the latter being such that they may be considered to be sufficiently small to serve the role of differentials in the following calculations, but, at the same time, sufficiently large for each to contain a large number dn of velocity-points. This, of course, means that the number of volume-elements per cubic centimetre is very much less than n.

To each of the volume-elements let a number of the series 1, 2, 3, . . . etc., be assigned, and let dn_1, dn_2, dn_3, . . . etc. represent the number of velocity-points that the volume-elements respectively contain. Then

$$n = dn_1 + dn_2 + dn_3 + \ldots$$

Now the velocities of the molecules whose velocity-points all lie within a given volume-element differ but slightly, and we shall, therefore, consider that the velocities of the molecules within a given element are the same. Corresponding to a given macroscopic state of a gas, i.e. a state determined by its pressure, volume, temperature, etc., will be a certain distribution of the velocity-points among the volume-elements. To this distribution will correspond a number of microscopic states or complexions as such states are sometimes termed. This means that we may interchange the individual molecules between the various volume-elements (the number in each remaining the same) without changing the given distribution. For example, if the number of volume-elements were equal to n, and there were one molecule per volume-element, the number of micro-states would be $n\,!$, i.e. the number of permutations of n things taken n at a time. As, however, there are dn_1, dn_2, etc., molecules whose velocities are the same (or lie within certain narrow limits) the number of micro-states is

$$w = \frac{n\,!}{dn_1\,!\,dn_2\,!\,dn_3\,!\,\ldots} \qquad (1\text{–}52)$$

and the value of w will depend on the number of velocity-points dn_1, dn_2, etc., there are in each volume-element.

When a substance is not in a state of macroscopic equilibrium we know that its state changes until equilibrium is reached. This change is, of course, brought about by a change in the distribution of the velocity-points and, of course, leads to a change in w. When equilibrium is attained $\delta w/\delta t = 0$, or w is a maximum. Hence a change from a state which is not one of equilibrium to one that is involves an increase in the number of micro-states. It follows that the distribution function we are seeking is that which renders (1–52) a maximum.

In order to transform (1–52), use is made of an approximation formula due to Stirling. We have

$$\log n! = \log 1 + \log 2 + \log 3 + \ldots \log n$$

and
$$\log n^n = \log n + \log n + \ldots \log n$$

so that
$$\log\ (n!/n^n) = \log \frac{1}{n} + \log \frac{2}{n} + \ldots \log \frac{n}{n}$$

Putting $1/n = dx$, we have

$$\frac{1}{n} \log (n!/n^n) = \log dx \,.\, dx + \log 2dx \,.\, dx + \ldots \log ndx \,.\, dx$$

$$= \int_0^1 \log x \,.\, dx = -1$$

Thus $\log n! = n \log n - n = \log (n/\varepsilon)^n$

so that
$$n! = (n/\varepsilon)^n$$

Substituting in (1–52)

$$w = \frac{(n/\varepsilon)^n}{(dn_1/\varepsilon)^{dn_1} (dn_2/\varepsilon)^{dn_2} (dn_3/\varepsilon)^{dn_3} \ldots}$$

$$= \frac{n^n}{dn_1{}^{dn_1} dn_2{}^{dn_2} dn_3{}^{dn_3} \ldots}$$

since
$$dn_1 + dn_2 + dn_3 + \ldots = n$$

Thus
$$\log w = - [dn_1 \log dn_1 + dn_2 \log dn_2 + dn_3 \log dn_3 + \ldots] + n \log n \quad . \quad . \quad (1–53)$$

From (1–51) we have

$$dn = Fd\tau$$

and substituting in (1–53)

$$\log w = - \Sigma Fd\tau \log (Fd\tau) + n \log n$$

$$= - \Sigma F \log Fd\tau - \Sigma Fd\tau \log d\tau + n \log n$$

Substituting dn for $Fd\tau$ and replacing the sign of summation by that of integration,

$$\log w = -\int F \log F d\tau + n \log n + A$$

where A is a constant of integration. Further simplifying,

$$\log w = -\int F \log F d\tau + A'$$

or, ignoring the additive constant, we may write

$$H = -\int F \log F d\tau \qquad (1\text{–}54)$$

where

$$H = \log w \qquad (1\text{–}55)$$

By (1–55) an expression has been obtained relating the distribution function to the number of micro-states, or probability, as w is sometimes termed.

Now in order to determine F we make use of the fact that F is subject to the following conditions or *constraints*

$$\int F d\tau = n \qquad (1\text{–}56)$$

$$\int \tfrac{1}{2} m(u^2 + v^2 + w^2) F d\tau = ET \qquad (1\text{–}57)$$

which express the facts that the number of molecules per cubic centimetre and the energy of the gas are constant. These two conditions, in conjunction with the fact that, in a steady state (1–54) must be an extremum, constitute an isoperimetric problem in the Calculus of Variations. Hence the value of F to make (1–54) an extremum is obtained from the solution of the Euler-Lagrange equation; i.e.

$$[\log_\varepsilon F + 1 + \lambda + \mu(u^2 + v^2 + w^2)] = 0 \qquad (1\text{–}58)$$

where λ and μ are constants. (1–58) leads to

$$F = B\varepsilon^{-b\,(u^2 + v^2 + w^2)} \qquad (1\text{–}59)$$

where B and b are constants. Thus it is found that the condition of equilibrium is the Maxwell Law of Distribution of Velocities. Comparison of (1–51) with (1–5) and (1–12) shows that we may write

$$F = nK_1\varepsilon^{-\frac{u^2+v^2+w^2}{a^2}} = nK_1\varepsilon^{-\frac{c^2}{a^2}} \quad . \quad . \quad (1\text{–}60)$$

where $K_1 = K^3$.

Referring to (1–59) we may further generalize matters by replacing the translational kinetic energy of the molecules by their total energy.* In this case we may put

$$F = B\varepsilon^{-2bE} \quad . \quad . \quad . \quad . \quad . \quad (1\text{–}61)$$

where $b = 1/ma^2$.

The Quantum Theory

For large-scale or macroscopic phenomena it is invariably found that these are governed by the laws of the Newtonian or classical system of dynamics. It was at one time assumed that the behaviour of matter in its finer or microscopic states was also explainable by this system. The behaviour of certain phenomena, however, is found to be inconsistent with the classical system, the first notable instance being the rate of energy radiation from a black body. In endeavouring to find the law relating the temperature of a hot body to the distribution of radiant energy amongst its emitted wavelengths, Planck introduced the quantum theory, one of the most important theories in physics. Since its introduction the quantum theory has found many applications, particularly to the finer states of matter, and hence it is found to have numerous applications in electronics. For example, the atom, photo-electricity, thermionic emission, the theory of spectra, ionization, X-rays, etc., are all governed by quantum considerations and thus attention must be given to this important theory.

The basic idea of the quantum theory is that energy, no less than matter, is atomic in nature. That is, energy is not absorbed or emitted continuously, but in small elements or quanta. These quanta are, as atoms of a given substance, all identical, and a given amount of energy contains an integral number of quanta. Just as atoms of different elements differ in magnitude, so also do quanta. For example, when energy exchanges occur between matter and radiation, a quantum of energy, E, in the form of radiation, is equal to hf, where h is a constant and f the frequency of vibration. Thus, in this case, the exchange takes the form of a monochromatic vibration of frequency f. Evidently the magnitude of a quantum is proportional to f. Conversely, when radiant energy is absorbed by matter, it is again quanta which are taken up.

---

* See p. 160.

As the phenomena of heat radiation were among the first to show the inadequacy of classical dynamics to deal with the fine-scale structure of matter and hence to introduce the quantum theory, such phenomena will now, briefly, be dealt with. Studies of heat radiation from black bodies reveal that the radiant energy is distributed among the emitted vibrations or wavelengths in a particular manner. The original experiments of Lummer and Pringsheim revealed that the energy distribution is such that the curve relating energy to wavelength (for a given black-body temperature) possesses a single maximum, the energy falling continually from this maximum and becoming zero for λ, the wavelength, being either zero or infinity. Fig. 1–4 shows some results due to these experimenters, from which it will be noted that the maximum shifts in the direction of shorter wavelengths as the temperature is increased.

Reference to Fig. 1–2 shows that there is, at least, a superficial resemblance between this curve and those of Fig. 1–4. This is a fact which was very early noticed, and led to attempts to find a connexion between the energy distribution among the wavelengths and the energy distribution among the molecules. In the case of a gas, as we have already seen, it is not possible to state the exact velocity possessed by any molecule. It is only possible to specify the number of molecules the speeds of which fall within a given speed range dc. Similarly with radiation, it is not possible to specify the amount of radiant energy at a given wavelength λ, but only an energy interval dE corresponding to a range $d\lambda$. Hence with black-body or thermal radiation it is not possible to state the energy possessed by a monochromatic radiation. However, in practice, a narrow waveband, usually 10^4 Ångstroms, is taken and regarded as monochromatic.

ENERGY DISTRIBUTION FORMULAE

Efforts were naturally made to formulate the relation between rate of energy radiation and wavelength for results such as those of Fig. 1–4. Reasoning from thermodynamics, Wien derived the general law

$$E_\lambda = K\lambda^{-5}f(\lambda T) \quad . \quad . \quad . \quad (1\text{–}62)$$

where E_λ is the emissive power (rate of energy radiation per unit range of wavelength) for a wavelength λ, $f(\lambda T)$ an unknown function of λT, and K a constant. In view of the fact that $f(\lambda T)$ is unknown, it is necessary to make assumptions concerning the mechanism of the radiator to elucidate f.

Making use of Maxwell's Law of Velocity Distribution, Wien deduced a formula for E_λ based on the following ideas and assumptions.

Let A be an enclosure the internal surface of which is totally reflecting, and B a perfectly transparent vessel within A. The space between A and B is assumed to be filled with a gas capable of absorbing and emitting radiations. The space within B will be filled with black-body radiation corresponding to the temperature of the gas. It is assumed that each molecule of the gas emits rays, the wavelength of which depends only on the velocity of the molecule, the intensity of the radiation being regarded as a function of the wavelength. As the wavelength is a function of the molecular velocity, the latter may be considered as a function of the wavelength. The energy of radiation of wavelength between λ and

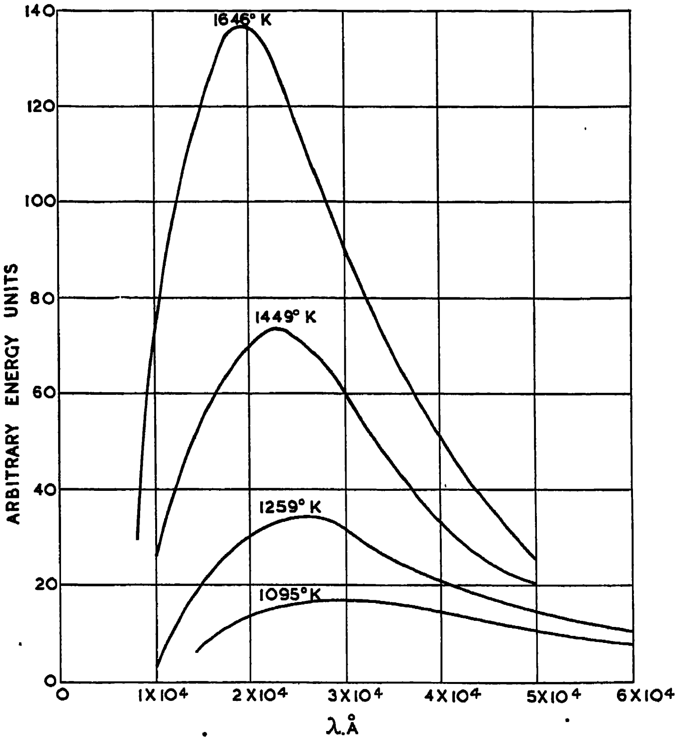


FIG. 1–4

$\lambda + d\lambda$ is proportional to the number of molecules sending out waves of this magnitude and to a function of the molecular velocity c.

Now according to the kinetic theory the number of molecules whose velocities lie between the limits c and $c + dc$ is (page 8)

$$Ac^2\varepsilon^{-\frac{c^2}{a^2}}dc \quad . \quad . \quad . \quad . \quad (1\text{–}63)$$

where $C^2 = \frac{3}{2}a^2$ and A is a constant. Expressing c as a function of λ and remembering that $C^2 \propto T$, (1–63) may be written

$$A_1 f_1(\lambda)\varepsilon^{-\frac{f(\lambda)}{T}}\,d\lambda$$

As the energy is proportional to this expression and some other function of λ, we obtain

$$E_\lambda = A_2 F(\lambda)\varepsilon^{-\frac{f(\lambda)}{T}} \quad . \quad . \quad . \quad (1\text{–}64)$$

where A_1 and A_2 are constants. Comparing (1–64) with (1–62) gives Wien's formula

$$E_\lambda = A_3\lambda^{-5}\varepsilon^{-\frac{b}{\lambda T}} \; . \quad . \quad . \quad . \quad (1\text{–}65)$$

where $F(\lambda) = \lambda^{-5}$, $f(\lambda) = 1/\lambda$, and $f(\lambda T) = \varepsilon^{-\frac{b}{\lambda T}}$

The assumptions made by Wien in arriving at (1–65) do not rest on a particularly firm foundation and hence it is not surprising that the formula only covers a limited part of the radiation spectrum. In particular, agreement is found for low values of λ and T.

We shall now turn to Planck's treatment of the subject, which, as previously stated, led to the introduction of the idea of quanta.

Let a dynamical system (such as a solid body) be considered composed of a very large number of identical parts or units, each unit being regarded as capable of executing vibrations. For example, in a solid body the units will be atoms the motion of which is restricted by the lattice structure. Assuming the atoms to be restricted by a quasi-elastic force, a displacement x from the mean position will call into existence a restoring force μx, where μ is a constant.

Now, according to the kinetic theory, the probability of a unit possessing velocities (u, v, w) and co-ordinates (x, y, z) within ranges du, dv, dw, and dx, dy, dz, respectively is, from (1–4) and (1–61),

$$\varepsilon^{-2bE}dudvdwdxdydz = \varepsilon^{-2bE}d\tau$$

where $d\tau$ is a volume-element in 6-dimensional space.* E is the total energy of a unit and, according to page 161, is half kinetic and half potential. Hence

---

* See p. 32.

$$E = \tfrac{1}{2}m(u^2 + v^2 + w^2) + \tfrac{1}{2}\mu(x^2 + y^2 + z^2)$$

and the number of units having this energy may be written

$$B\varepsilon^{-2bE}d\tau$$

where the positional co-ordinates lie between $x + dx$, $y + dy$, and $z + dz$. If E is a given amount of energy the numbers of units having energies, O, E, $2E$, $3E$, etc., will stand in the ratio

$$1 : \varepsilon^{-2bE} : \varepsilon^{-4bE} : \varepsilon^{-6bE} : \ldots$$

so that if A have zero energy, the numbers having energies E, $2E$, $3E$, . . . will be $A\varepsilon^{-2bE}$, $A\varepsilon^{-4bE}$, $A\varepsilon^{-6bE}$, etc. If M is the number of vibrations produced by the system, then the energies of these vibrations will be equal to the various values E, $2E$, $3E$, etc. Also

$$\begin{aligned} M &= A(1 + \varepsilon^{-2bE} + \varepsilon^{-4bE} + \varepsilon^{-6bE} + \ldots) \\ &= \frac{A}{1 - \varepsilon^{-2bE}} \quad \ldots \quad (1\text{–}66) \end{aligned}$$

The total energy of the vibrations is

$$\begin{aligned} &EA\varepsilon^{-2bE} + 2EA\varepsilon^{-4bE} + 3EA\varepsilon^{-6bE} + \ldots \\ &= EA\varepsilon^{-2bE}(1 + 2\varepsilon^{-2bE} + 3\varepsilon^{-4bE} + \ldots) \\ &= \frac{EA\varepsilon^{-2bE}}{(1 - \varepsilon^{-2bE})^2} \end{aligned}$$

or from (1–66)

$$\frac{ME}{\varepsilon^{2bE} - 1} \quad \ldots \quad (1\text{–}67)$$

Now if the vibrations are those of wavelengths lying between λ and $\lambda + d\lambda$, then it may be shown* that the value of M is $8\pi\lambda^{-4}d\lambda$ and thus the vibrational energy is

$$8\pi\lambda^{-4}\frac{E}{\varepsilon^{2bE} - 1}\,d\lambda$$

or, substituting for b,

$$\frac{8\pi\lambda^{-4}Ed\lambda}{\varepsilon^{\frac{E}{kT}} - 1} \quad \ldots \quad (1\text{–}68)$$

---

* *Theory of Heat*, T. Preston, p. 290 (Macmillan).

To determine the value of E we shall consider a unit of the system under the influence of a restoring force μx. If m is the mass of a unit then

$$m\dot{x} + \mu x = 0$$

A solution of this equation is

$$x = A \sin \sqrt{\frac{\mu}{m}}\, t$$

where A is the amplitude of the oscillations, the oscillation frequency being given by

$$f = \frac{1}{2\pi}\sqrt{\frac{\mu}{m}}$$

The energy E of the oscillator is

$$E = \tfrac{1}{2}m\dot{x}^2 + \tfrac{1}{2}\mu x^2 + E_0 \quad . \quad . \quad . \quad (1\text{–}69)$$

where E_0 is any energy it may have when at rest in its equilibrium position. If we assume E_0 to be zero and put $m\dot{x} = y$, we may write

$$\frac{y^2}{2mE} + \frac{\mu x^2}{2E} = 1$$

which is the equation to an ellipse. Thus the representative point of the oscillator moves round the ellipse, making the circuit f times per second. The area of the ellipse is

$$2\pi E\sqrt{\frac{m}{\mu}} = E/f \quad . \quad . \quad . \quad (1\text{–}70)$$

Putting $E/f = h$, $E = hf$, and thus we see that the possible values of the energy of an oscillator are given by nhf, where $n = 1, 2, 3$, etc. It is evident that the representative point of the oscillator divides the xy plane into equal areas of value h. This quantity is known as Planck's Constant and has the value $6{\cdot}554 \times 10^{-27}$ erg.-sec. Substituting for E in (1–68) we have

$$\frac{8\pi\lambda^{-4}hf\,d\lambda}{\varepsilon^{\frac{hf}{kT}} - 1} \quad . \quad . \quad . \quad . \quad (1\text{–}71)$$

or, writing c/λ for f

$$E_\lambda d\lambda = \frac{8\pi hc\lambda^{-5}d\lambda}{\varepsilon^{\frac{hc}{kT\lambda}} - 1} \quad . \quad . \quad . \quad (1\text{–}72)$$

where E_λ may be defined as the rate of energy radiation per unit range of wavelength at wavelength λ and c is the velocity of light.

The result given by (1–72) is Planck's formula for the energy distribution of a black-body or temperature radiator and is found to be in close agreement with experimental results. It appears that temperature radiators must be regarded as composed of numerous oscillators capable of vibrating with every conceivable frequency. Each possesses an integral number of quanta of energy, a quantum being equal to hf. Thus an oscillator of frequency f is capable of producing monochromatic wave trains of radiation, the radiant energy leaving the oscillator discontinuously in bursts or quanta of magnitude hf.

Planck's Radiation Law contains all other laws of radiation. For example, for low values of λ and T, $\varepsilon^{\frac{hc}{kT\lambda}}$ is large compared with unity, in which case (1–72) becomes

$$E_\lambda d\lambda = 8\pi hc\lambda^{-5}\varepsilon^{-\frac{hc}{kT\lambda}}\,d\lambda$$

which is of the same form as Wien's Law expressed by (1–65).

Wien's Displacement Law

This law states that $\lambda_m T$ = constant where λ_m is the wavelength at which maximum energy emission occurs from a temperature radiator. Differentiating E_λ with respect to λ and equating to zero, (1–72) becomes

$$\frac{-8\pi hc\left[-\frac{\lambda^5 hc}{kT\lambda^2}\varepsilon^{hc/kT\lambda} + 5\lambda^4(\varepsilon^{hc/kT\lambda} - 1)\right]}{\lambda^{10}(\varepsilon^{hc/kT\lambda} - 1)^2} = 0$$

or

$$\frac{\lambda^4 hc}{kT\lambda}\varepsilon^{hc/kT\lambda} = 5\lambda^4(\varepsilon^{hc/kT\lambda} - 1)$$

Putting $hc/kT\lambda = y$, we have

$$\frac{y\varepsilon^y}{\varepsilon^y - 1} = 5$$

It is evident that there is a root in the neighbourhood of 5. The exact value is 4·965. Hence

$$\frac{hc}{kT\lambda_m} = 4{\cdot}965$$

and

$$\lambda_m T = \frac{hc}{4{\cdot}965k} = \text{constant} \quad . \quad . \quad . \quad (1\text{–}73)$$

which is Wien's Displacement Law.

If we write $c/f = \lambda$ and substitute in (1–72) we obtain

$$E_f = \frac{8\pi h f^3}{c^3(\varepsilon^{hf/kT} - 1)}$$

Putting $hf/kT = x$,

$$E_f = \frac{8\pi k^3 T^3}{c^3 h^2} \cdot \frac{x^3}{\varepsilon^x - 1}$$

The total energy is obtained by integrating over all frequencies from 0 to ∞. Thus

$$E_\tau = \frac{8\pi h}{c^3} \int_0^\infty \frac{f^3 df}{\varepsilon^{hf/kT} - 1} = \frac{8\pi k^4 T^4}{c^3 h^3} \int_0^\infty \frac{x^3 dx}{\varepsilon^x - 1}$$

Now

$$\int_0^\infty \frac{x^3 dx}{\varepsilon^x - 1} = \int_0^\infty (\varepsilon^{-x} + \varepsilon^{-2x} + \varepsilon^{-3x} + \ldots)x^3 dx \, .$$

$$= 6\left(1 + \frac{1}{2^4} + \frac{1}{3^4} + \ldots\right) = \frac{\pi^4}{15}$$

Hence

$$E_\tau = \frac{8\pi^5 k^4}{15 c^3 h^3} T^4 = \sigma T^4 \quad . \quad . \quad . \quad . \quad (1\text{–}74)$$

which is the Stefan-Boltzmann Law, this law stating that the total energy emitted by a black body varies as the fourth power of the absolute temperature.

From the two equations given by (1–73) and (1–74) and experimental measurements of E and λ_m the values of h and k may be derived. Thus h is found to be $6{\cdot}54 \times 10^{-27}$ erg.-sec. and k, $1{\cdot}37 \times 10^{-16}$ erg./degree.

QUANTIZATION

Having introduced the quantum theory in the manner in which it first made its appearance, i.e. in connexion with thermal radiation, it is now necessary to consider this highly important theory at greater length for its applications to the subject of electronics are numerous. The state of the elementary oscillators in the radiation theory may be said to be quantized and it is this particular aspect, i.e. quantization, which it is desired to pursue. As the position and motion of a body may be described in any one of several co-ordinate systems (such as cartesian or polar co-ordinates, for example) it is desirable to develop the quantization process in terms of what are known as *generalized co-ordinates.*

For the purpose of explaining such co-ordinates the position and velocity of a mass-point will be considered. Such a point is an idealized one, having no extension in space, but an intensely localized mass. In rectangular or cartesian co-ordinates the position of a point is given by its projections x, y, and z on three mutually perpendicular axes. In polar co-ordinates the position is given in terms of a radius vector, ρ, and two angular co-ordinates θ and ϕ. In addition to these two co-ordinate systems others exist, such as cylindrical and spherical co-ordinates. In view of these various systems it is found desirable to designate position co-ordinates in a generalized manner, as q_1, q_2, q_3, etc., there being as many co-ordinates at a point as the number of "dimensions" in which the point exists. Of course, in ordinary three-dimensional space a body has three co-ordinates only.

In describing the velocity of a body this is effected by giving its velocity components along the co-ordinate axes employed. However, because at high velocities the mass of a body varies with its velocity, and force $= d(mv)dt$ rather than mdv/dt, it is desirable to express the state of motion of a body in terms of momentum rather than velocity. Generalizing the momentum co-ordinates, we may write p_1, p_2, p_3, etc., in place of $(mv)_x$, $(mv)_y$, $(mv)_z$, where the latter three express the momentum components of a body in rectangular co-ordinates. Thus, in three-dimensional space, the co-ordinates q_1, q_2, q_3, p_1, p_2, p_3 are capable of completely describing the state of a point at any instant. In a general manner the kth co-ordinate is written q_k or p_k. We may now find general relations between the position and momentum co-ordinates, etc. Thus, when the mass of a body is independent of its velocity, we have

$$p_k = m\dot{q}_k$$

$$f_k = \dot{p}_k$$

where $f =$ force. Now the potential energy of a body may be written

$$\int f_k \delta q_k = E_p$$

and

$$f_k = \dot{p}_k = \frac{\delta E_p}{\delta q_k}$$

where E_p is the potential energy at the position under consideration. The kinetic energy of a mass-point is given by

$$E_k = \tfrac{1}{2}m(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2 + \text{etc.}) = \frac{p_1^2 + p_2^2 + p_3^2}{2m} \text{ etc.}$$

Differentiating,

$$\frac{\delta E_k}{\delta \dot{q}_k} = m\dot{q}_k$$

but $m\dot{q}_k = p_k$, so that

$$p_k = \frac{\delta E_k}{\delta \dot{q}_k}$$

As stated above, with the appropriate number of position and momentum co-ordinates the state of a mass-point, or any system of mass-points, may be described. The instantaneous state of such a system is known as the *phase* of that system. If the state of a body may be determined by three-position and three-momentum co-ordinates it may be imagined to exist in a hypothetical single six-dimensional space. This space is termed a *phase space* and the path of a body in such space is known as a *phase path.* In the case of a particle having two co-ordinates only, p_1, q_1, the path of the particle lies in a phase known as the phase plane. Now in determining the value of E in (1–69) we considered a particle under the influence of an elastic restoring force μx. The particle was found to execute vibrations in simple harmonic motion and hence constituted a simple oscillator. The representative point of the oscillator was found to move round an ellipse dividing the xy plane into equal areas, each of value h. The area enclosed by any curve in this plane is given by $A + nh$ where the inclusion of A indicates that E_0 in (1–69) is different from zero. In generalized co-ordinates we must replace x by q and y by p and in this case the quantum condition is given by

$$\int pdq = A + nh \text{ or } nh \text{ if } E_0 \text{ and } A = 0$$

the integration being performed over a cycle of movement.

Now the dimensions of p are $MV = MLT^{-1}$, and those of q are L; therefore $pq = ML^2T^{-1}$; these dimensions being those of energy × time or action. It thus appears that the phase paths of a body having a quantized motion divide the phase plane into areas which differ from each other by an integral number of units of action. The natural unit of action appears to be h, i.e. Planck's constant, although it is not evident why in nature action should be atomic in character. However, the same reflexion is possible of energy and matter. In reality such units are forced upon us as a consequence of experience.

We may now proceed to generalize the quantum conditions for a body having one degree of freedom and executing periodic motion. The area enclosed by the phase curve of the body is given by

$$J = \int\int dp dq$$

or, integrating, over p,

$$J = \int p dq = \int f(q) dq$$

this being known as the *phase integral*. The quantum condition that the action of a body may only change by an integral number of units of action may be written

$$\Delta J = h$$
$$J = J_0 + nh$$

where J_0 corresponds to E and A above. If $J_0 = 0$, then

$$J = \int p dq = nh \quad . \quad . \quad . \quad (1\text{–}75)$$

These statements are quite general and express the quantum restrictions that can be applied to any periodic system the properties of which are expressible in generalized co-ordinates.

The Atomic Theory

Various atomic models have been proposed the purpose of which is to explain the properties of matter and electrical phenomena. Such models have this in common that they consist of an assemblage of positive and negative charges of such magnitude and arrangement that the atom and matter are normally electrically neutral. Electrons are obviously constituents of the atom and, therefore, a positive charge, or charges, must exist whose magnitude is equal but opposite to that possessed by the total number of electrons within an atom. A system of point charges at rest is fundamentally unstable and to overcome this difficulty Rutherford suggested an atomic model in which the positive charges form an assembly at rest, termed the nucleus, and around which certain electrons revolve, the whole forming an arrangement similar to our solar system. As with the latter, stability is secured if the planetary electrons revolve at a velocity such that their centrifugal force balances that due to the electrostatic attraction of the nucleus. Assuming an electron to be traversing a circular orbit of radius r with velocity v, it will have an acceleration v^2/r, and, according to classical electrodynamical theory, will radiate energy continuously. To supply this energy the electron must constantly approach the nucleus and thus ultimately the atomic structure would collapse. It is by no means certain, however, that Newtonian dynamics are completely applicable to atomic physics and it will be shown later that the Rutherford model has been retained by an application to the atom of the quantum theory.

One of the earliest models of the atom was due to J. J. Thomson, who put forward the idea that the atom might consist of a sphere of positive electricity with electrons embedded within it. The diameter of this sphere was assumed from the kinetic theory to be of the order of 10^{-8} cm. An electron within the sphere at a distance r from the centre will be unaffected by the positive charge beyond r in accordance with the principle established by Faraday's ice-pail experiment. The charge within r is $\frac{4}{3}\pi r^3\rho$ where ρ is the space charge density of positive electrification. Assuming the Coulomb law of force, the force on the electron is

$$f = \tfrac{4}{3}\pi r^3\rho e/r^2 = \tfrac{4}{3}\pi r\rho e$$

which varies directly as the distance from the centre of the sphere. With atoms possessing more than one electron it was assumed that the electrons would remain at rest under the influence of the central attracting force and their mutual repulsions.

The Thomson atom was abandoned because of the nuclear scattering experiments of Rutherford and their results. By bombarding the atom with alpha particles it was found that the angles through which the latter were deflected were too large to be explained by a large diffuse positive sphere of the order of 10^{-8} cm. diameter. It was decided that the large-scale deflexions could only be accounted for if the positive charge were concentrated in a volume far less than that of a sphere of 10^{-8} cm. diameter. Thus the nuclear atom, already described, came into existence and experiments and calculations showed the diameter of the nucleus to be of the order of 10^{-12} cm. as against 10^{-8} cm. for the atom. Furthermore, it was established that the nucleus consists of a positive charge of Ze units (where Z is the atomic number of the atom) with Z electrons circulating around it. The nucleus is surrounded by a Coulomb field of force, i.e. $f \propto 1/r^2$, to within 10^{-12} cm. of the nucleus and is scarcely affected by the electrons which are at relatively large distances, 10^{-8} cm., from the atomic centre. As already stated, to balance the force of the nucleus it was necessary to assume that the electrons circulate about this. However, this raises the difficulty of continual energy radiation by the electron. To understand how this difficulty was met we must now consider Bohr's treatment of the nuclear atom.

THE RUTHERFORD-BOHR ATOM

It has already been shown that, according to the quantum theory, energy is not radiated continuously. By applying this principle to the atom, Bohr was able to retain the Rutherford model. Bohr's first postulate is that on certain permissible orbits the electron does not radiate energy as demanded by the classical theory. No other orbits than these are permissible, energy being either radiated or absorbed when an electron "jumps" from one such orbit to another. The second postulate states that when an electron passes from one orbit to another the energy emitted or absorbed is equal to hf. This means that should emission occur it will consist of a monochromatic radiation of frequency f. No explanation is given of the mechanism by which an electron passes between two orbits. Also there is apparently no connexion between the frequency of the radiation and the frequency of revolution of the electron in its orbit. As previously shown, the dimensions of Planck's constant h are energy × time or momentum × distance. These products are known as an "action." Bohr assumed that an orbit would be non-radiating or "stationary" if an electron in it possessed an amount of action equal to a multiple of h. For an electron describing a circular orbit we have on applying the phase integral (1–75)

$$\int_0^{2\pi} p\,dq = nh$$

where $p = mv$ and $dq = 2\pi dr$. As m and v are constant, integration gives

$$mv\,.\,2\pi r = nh \qquad . \quad . \quad . \quad (1\text{–}76)$$

where r is the radius of the orbit m, v the mass and velocity of the electron, and n an integer known as the quantum number of the orbit. If w is the angular velocity of the electron, then (1–76) may be written

$$2\pi mwr^2 = nh \qquad . \quad . \quad . \quad (1\text{–}77)$$

Bohr's two postulates will now be applied to the hydrogen atom. The latter consists of a positive nucleus, having a charge equal in magnitude to that of an electron, with a single electron revolving on an orbit which we may take to be circular. The quantum condition is

$$2\pi mwr^2 = nh \qquad . \quad . \quad . \quad (1\text{–}78)$$

and the condition of mechanical equilibrium

$$\frac{e^2}{r^2} = \frac{mv^2}{r} = mrw^2 \quad . \quad . \quad . \quad (1\text{–}79)$$

where e is the electronic charge.

Solving for r and w we obtain

$$r = \frac{n^2h^2}{4\pi^2me^2} \quad . \quad . \quad . \quad . \quad (1\text{–}80)$$

and

$$w = \frac{8\pi^3me^4}{n^3h^3} \quad . \quad . \quad . \quad . \quad (1\text{–}81)$$

For cases where the nuclear charge is greater than e, r and w may be written

$$r = \frac{n^2h^2}{4\pi^2mZe^2}$$

$$w = \frac{8\pi^3mZ^2e^4}{n^3h^3}$$

where Z is the number of charges forming the nucleus.

By substituting the numbers 1, 2, 3, etc., in (1–80) the radii of the possible stationary orbits are obtained. For $n = 1$, substituting the known values of the various quantities in (1–80) gives

$$r = 0{\cdot}52 \times 10^{-8} \text{ cm.}$$

a value in remarkable agreement with the radius of the hydrogen atom derived from the kinetic theory of gases. The values of w, derived from (1–81), correspond to the frequencies of ultra-violet light; although the radiated frequencies have apparently no connexion with w, it appears significant that they are of the same order of magnitude. In order to find the radiated actual frequencies on the basis of the second Bohr postulate, the energy must be found in the nth quantum state.

The kinetic energy of the electron is $\frac{1}{2}mv^2$ or, from (1–79), $e^2/2r$. As the force between nucleus and electron is attractive, the potential energy will be a maximum when the electron is at an infinite distance from the nucleus. At a distance r the potential energy is $-\, e^2/r$ and hence the total energy is

$$\frac{e^2}{2r} - \frac{e^2}{r} = -\frac{e^2}{2r} \quad . \quad . \quad . \quad (1\text{–}82)$$

Substituting for r from (1–80), (1–82) becomes

$$-\frac{2\pi^2me^4}{n^2h^2} \quad . \quad . \quad . \quad . \quad (1\text{–}83)$$

Now, according to the second postulate, the energy emitted or absorbed in the form of radiation when an electron changes from one orbit to another is

$$E_i - E_f = hf$$

from which $$f = \frac{E_i - E_f}{h}$$

From (1–83)

$$f = \frac{2\pi^2 me^4}{h^3}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \quad . \quad . \quad . \quad (1\text{–}84)$$

Where n_i and n_f respectively correspond to E_i and E_f; i.e. the energies in the initial and final states of an energy change. Thus the lines in the hydrogen spectrum should be obtainable by substituting different integers for n_i and n_f in (1–84). n_f is the quantum number of the orbit into which the electron falls, while n_i is that of the orbit from which it starts.

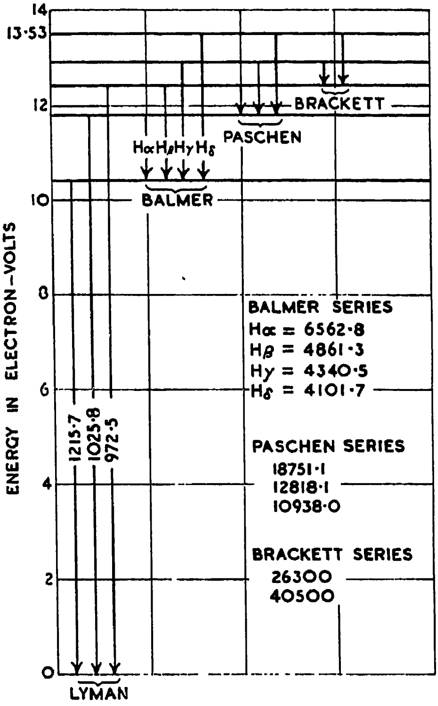


Fig. 1–5

Actually it is found that (1–84) does represent all the spectral lines emitted by hydrogen. The sequence of lines with $n_f = 2$, $n_i = 3$, 4, 5, . . . , is known as Balmer's series, i.e. a series for which Balmer originally found an empirical formula, similar in form to that of (1–84), in 1885. The lines crowd together as the current number n_i increases, and at the limit $n_i = \infty$ there is an accumulation point of spectral lines. The intensity of the lines decreases as the limit is approached. For $n_f = 1$, $n_i = 3$, 4, 5, . . . we have a series known as the Lyman series lying in the far ultra-violet region. The lines corresponding to $n_f = 3$, $n_i = 4$, 5, . . . , and to $n_f = 4$, $n_i = 5$, 6, 7, . . . , are in the infra-red and are, respectively, termed the Paschen and Brackett series after their discoverers.

In order to visualize better the effects of electron transitions within the atom we may employ an energy level diagram shown by Fig. 1–5. The horizontal lines in this represent the various orbits or energy levels within the atom, the direction of the arrows indicating the direction in which the electron transitions take place when radiation occurs. For absorption the arrow heads would, of course, be reversed. The left-hand ordinate of the diagram is in electron-volts, and the energy emitted in the form of radiation is the difference between the initial and final levels. The abscissa corresponds to the innermost orbit of the atom, which is known as the "ground state."

In spectroscopy it is customary to use the wave number instead of the frequency f. This number is the number of waves per centimetre and Balmer's empirical formula gave this quantity. The formula is written

$$\gamma = R_{\mathrm{H}}\left(\frac{1}{4} - \frac{1}{n^2}\right) . \qquad . \qquad . \qquad . \qquad (1\text{–}85)$$

where R_{H} is known as the Rydberg constant for hydrogen, its value being $1{\cdot}09677 \times 10^5$ cm.$^{-1}$ If $n_f = 2$ in (1–84) it will be noted that (1–84) and (1–85) are identical in form. Dividing (1–84) by c (the velocity of light) to convert frequency to wave number we have for the Rydberg constant

$$R_{\mathrm{H}} = \frac{2\pi^2 m e^4}{ch^3} \qquad . \qquad . \qquad . \qquad . \qquad . \qquad (1\text{–}86)$$

$$= 1{\cdot}09737 \times 10^5 \text{ cm.}^{-1}$$

an extremely close agreement with the empirical figure.

THE MOTION OF THE NUCLEUS

In deriving the foregoing results it has been assumed that the nucleus is at rest, a condition only true for infinitely heavy nuclei. Actually the electron and nucleus revolve round their common centre of gravity. M is the mass of the nucleus, r_2 its distance from the centre of gravity of the system, and r_1 that of the electron, then

$$mr_1 = Mr_2 \qquad . \qquad . \qquad . \qquad . \qquad (1\text{–}87)$$

The condition of mechanical equilibrium for the nucleus is

$$\frac{e^2}{(r_1 + r_2)^2} = Mr_2 w^2 \qquad . \qquad . \qquad . \qquad . \qquad (1\text{–}88)$$

and for the electron

$$\frac{e^2}{(r_1 + r_2)^2} = mr_1w^2 \quad . \quad . \quad . \quad . \quad (1\text{–}89)$$

Now $(r_1 + r_2) = r$, and hence (1–89) becomes

$$\frac{e^2}{(r_1 + r_2)^2} = \frac{mM}{m + M} rw^2 \quad . \quad . \quad . \quad (1\text{–}90)$$

The total energy of the system is

$$\begin{aligned} &\tfrac{1}{2}Mr_2^2w^2 + \tfrac{1}{2}mr_1^2w^2 - \frac{e^2}{r_1 + r_2} \\ &= \frac{w^2}{2}(Mr_2^2 + mr_1^2) - \frac{e^2}{r_1 + r_2} \\ &= \frac{w^2}{2} mr_1(r_1 + r_2) - \frac{e^2}{r_1 + r_2} \\ &= \frac{w^2}{2} \cdot \frac{mM}{m + M}(r_1 + r_2)^2 - \frac{e^2}{r_1 + r_2} \quad . \quad . \quad (1\text{–}91) \end{aligned}$$

Comparing (1–79) and (1–82) with (1–90) and (1–91) we see that when the nucleus is taken into account (1–79) and (1–83) must be multiplied by a factor $M/(m + M)$. Thus, the corrected formula for the frequency is

$$f = 2\pi^2 \frac{mM}{m + M} \frac{e^4}{h^3} \left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \quad . \quad . \quad (1\text{–}92)$$

IONIZED HELIUM

The spectrum of the hydrogen atom is the only one for which calculations can be normally made by Bohr's theory. For atoms possessing two or more extra-nuclear electrons the difficulties become insuperable. If, however, an atom can be stripped of these electrons, then the return of the first electron produces a state which may receive mathematical treatment. A case in point is that of helium. The nuclear charge of this atom is twice that of hydrogen, and if one of its two extra-nuclear electrons can be removed, a structure exists for which the spectrum lines may be deduced. Such atoms can be produced and the spectrum should be given by

$$f = 4 \cdot \frac{2\pi^2me^4}{h^3} \left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \cdot \quad . \quad . \quad (1\text{–}93)$$

the factor 4 occurring due to the double charge on the nucleus. We thus see that the frequencies of those lines corresponding to the hydrogen lines are four times as great as those of the hydrogen lines. Hence the helium spectrum corresponding to the Lyman and Balmer series lies in the far ultra-violet.

Before the deduction of (1–84) from Bohr's theory, Pickering and Fowler had discovered a spectrum which could be represented by

$$4R\left(\frac{1}{n_f^2}-\frac{1}{n_i^2}\right)$$

where R differed but slightly from the Rydberg constant for hydrogen. This difference is now known to be due to the differences in the motions of the hydrogen and helium nuclei. Slight though the difference is, it may be determined spectroscopically and from the determination the ratio of the mass of an electron to that of a proton (hydrogen nucleus) found. Putting R_{He} for the Rydberg constant for helium, and M_{He} and M_{H} for the respective nuclear masses, we have from (1–92)

$$\begin{aligned}\frac{R_{\mathrm{H}}}{R_{\mathrm{He}}} &= \frac{M_{\mathrm{H}}}{m+M_{\mathrm{H}}}\bigg/\frac{M_{\mathrm{He}}}{m+M_{\mathrm{He}}}\\ &= \frac{1}{1+x}\Bigg/\frac{M_{\mathrm{He}}/M_{\mathrm{H}}}{x+\dfrac{M_{\mathrm{He}}}{M_{\mathrm{H}}}}\\ &= \frac{1+\dfrac{M_{\mathrm{H}}}{M_{\mathrm{He}}}x}{1+x}\end{aligned}$$

from which
$$x = \frac{\dfrac{R_{\mathrm{H}}}{R_{\mathrm{He}}}-1}{\dfrac{M_{\mathrm{H}}}{M_{\mathrm{He}}}-\dfrac{R_{\mathrm{H}}}{R_{\mathrm{He}}}}$$

where $x = m/M_{\mathrm{H}}$.

Now
$$R_{\mathrm{H}} = 109677 \qquad M_{\mathrm{H}} = 1{\cdot}0072$$
$$R_{\mathrm{He}} = 109722 \qquad M_{\mathrm{He}} = 4{\cdot}0016$$

whence
$$\frac{m}{M_{\mathrm{H}}} = \frac{1}{1838}$$

Thus the ratio of the mass of an electron to that of the hydrogen *atom* is 1/1839.

It will be appreciated that the Bohr atom with its somewhat revolutionary assumptions is largely justified by its striking success in giving the diameter of the hydrogen atom and the frequencies of the spectral lines. Although the Bohr model cannot be considered as final it has undoubtedly thrown considerable light on atomic structure. In particular it seems certain that in all atoms the electrons are arranged in different energy levels, and that absorption or emission of radiant energy is effected by a change of level governed by quantum laws. When the electron of the hydrogen atom is on the innermost orbit the total energy is a minimum and, from (1–83), is equal to

$$-\frac{2\pi^2 mMe^4}{n^2h^2(m+M)} \quad . \quad . \quad . \quad (1\text{–}94)$$

This state is the normal condition for the atom and is known as the ground state. If the electron is completely removed from the atom, i.e. $n = \infty$, the total energy is zero and the potential energy a maximum. The complete removal of an electron, however, leaves the atom with a positive charge and in this condition the atom is said to be ionized and is termed an ion. In any state intermediate between the ground state and ionization the atom is said to be excited. These remarks, of course, apply to all atoms, whether their structure is similar to that of the hydrogen atom or not.

Now to raise an atom from the ground state to some higher state involves the absorption of energy by the atom. If E_i is the energy in the ground state and E_f the energy after absorption, then E_f corresponds to a state of excitation and the energy available for emission on a return to normal conditions is $E_f - E_i$. Under certain conditions this energy is given up in the form of monochromatic radiation in accordance with the quantum condition

$$hf = E_f - E_i$$

The transition is generally extremely rapid, occurring in approximately 10^{-8} sec. A method of inducing excitation is by bombarding the atom with electrons, although this is not the only available method. The kinetic energy of an electron is $\frac{1}{2}mv^2$ and if this is equal to $E_f - E_i$ then, by collision, the electron may raise the energy state of the atom from E_i to E_f. Hence, for excitation to occur, we must have

$$\tfrac{1}{2}mv^2 \geq E_f - E_i$$

An electron which has experienced a free fall through a potential difference V has energy Ve. If V_c is the potential difference necessary to give the electron sufficient energy to produce excitation, then

$$eV_c = E_f - E_i$$

and V_c is known as an excitation potential. If V is sufficiently high to give the electron the energy necessary to remove completely an electron from within the atom, then V_i is known as the ionization potential. From (1–94) the energy required for ionization of the hydrogen atom should be

$$\frac{2\pi^2 mMe^4}{n^2h^2(m+M)} \quad . \quad . \quad . \quad . \quad (1–95)$$

MEASUREMENT OF EXCITATION POTENTIALS

In order to measure excitation potentials the apparatus of Fig. 1–6 may be employed. This consists of a heated cathode C (which serves as a source of electrons), a grid G, and a plate A. The grid is at a positive potential with respect to the cathode, while the plate is at a negative potential with respect to the grid. The electrode system is enclosed in a glass tube and exhausted to the degree that the mean free path is greater than the cathode-grid distance, but less than the cathode-plate distance. The electrons are accelerated in the cathode-grid space, the majority arriving at the grid without collision. The energy of these electrons is

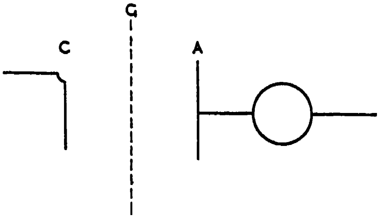


FIG. 1–6

$$eV_{cg} = \tfrac{1}{2}mv^2 \quad . \quad . \quad . \quad (1–96)$$

where V_{cg} is the cathode-grid potential difference. If this energy is sufficient to produce ionization, positive ions will be formed in the grid-anode space and will be directed to the anode by the field due to the grid-anode potential difference. Thus a suitable instrument connected in the cathode-anode circuit will indicate a current as soon as V_{cg} is such that ionization occurs.

A current in the anode circuit is, however, not necessarily due to ionization. If the electron energy is less than that needed for ionization, but sufficient for excitation, excited atoms will result. When these return to the ground state they may produce ultra-violet radiation which acting on the anode will produce electrons by photo-excitation. These electrons will pass to the grid and cause an indication on the meter in the grid-anode circuit in the same direction as that produced by positive ions flowing to the anode. Hence the method of Fig. 1–6 does not enable a determination to be made as to whether a sudden appearance of current in the grid-anode circuit is due to excitation or ionization.

The nature of the characteristic obtained as V_{cg} is raised is shown by Fig. 1–7, which concerns a hydrogen-filled tube. It will be noted that no current is indicated until V_{cg} is about 10 volts and that this current increases linearly up to about 16 volts. At this point a break occurs in the curve and the current increases more rapidly. Thus it may be said that the method of Fig. 1–6 indicates that hydrogen possesses critical potentials at approximately 10 and 16 volts.

The previously described method of determining critical potentials is not capable of great accuracy, apart from its inability to differentiate between excitation and ionization potentials. This is because of (1) the difficulty of accurately fixing a sudden break in a rapidly rising curve, (2) the contact difference of potential between grid and cathode, and (3) the initial velocity of emission of electrons from the cathode. A method possessing greater accuracy is the following. The gas pressure in the tube is increased until the mean free path of the electrons is smaller than the cathode-grid distance. The grid is placed just in front of the anode, the potential of the latter being about half a volt lower than that of the grid. Electrons passing through the grid must, therefore, possess sufficient energy to overcome this reverse field if they are to be collected by the anode. During their passage across the cathode-grid space, the electrons make numerous collisions with neutral gas molecules. However, providing the molecules have no electron affinity, and V_{cg} is less than the lowest critical potential, the collisions will be elastic and a number of electrons will arrive at the grid with energy corresponding to V_{cg}. If this is equal to or greater than half a volt the electrons will be collected by the anode and a current will be indicated in the anode circuit. As V_{cg} is increased, the number of electrons reaching the anode increases with a consequent increase in anode current. The latter will continually increase with V_{cg} until a critical potential is reached. At this juncture inelastic collisions will occur, the result of this being that electrons will give up their entire energy to atoms to produce a state of excitation in the latter. Such electrons will, of course, now be without the necessary energy to penetrate the

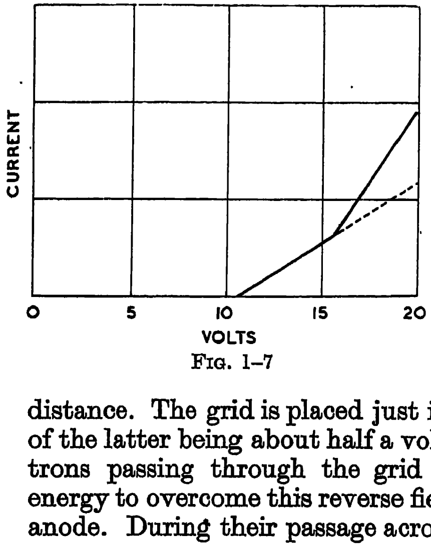


FIG. 1–7

grid, and reach the anode, with the result that the anode current will commence to fall. Thus the voltage for which the current is a maximum corresponds to a critical potential for the gas. As V_{cg} is continually increased, the inelastic collisions occur nearer the cathode and ultimately at such a distance from the grid that the energy subsequently acquired by the electrons is sufficient to carry them through the grid to the anode. At this stage the anode current passes through a minimum and again rises.

Now if V_{cg} is increased to approximately twice the critical potential, the inelastic collisions will occur at about half-way between cathode and grid. During the remaining half of the cathode-grid distance the electrons will again acquire sufficient energy for an inelastic collision which will, therefore, occur in the vicinity of the grid. Due to this, the current will again fall and it is apparent that if V_{cg} is continually increased the current/voltage curve will be characterized by a series of maxima and minima, the difference between two adjacent maxima giving the critical potential. Should the gas possess a number of critical potentials each will produce its own set of maxima.

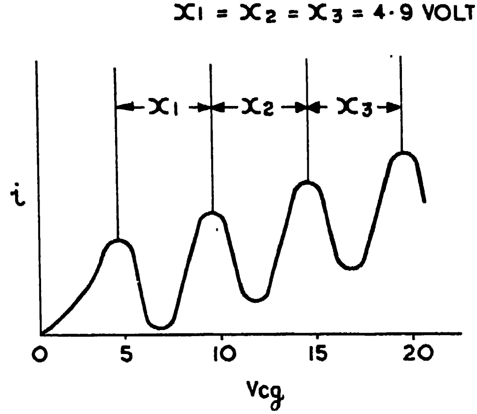


FIG. 1–8

A typical characteristic for mercury vapour is shown in Fig. 1–8, in which it will be noted that the distance to the first maximum is less than between succeeding ones. This is due to the cathode-grid contact potential and the initial velocity of electron emission. It will be noted that, for the case shown, these effects amount to 0·8 volt.

By means of the foregoing and similar methods, the excitation and ionization potentials of various atoms and molecules have been determined. The results give valuable information regarding the structure of the atom and confirmation of the validity of the application of quantum laws. We see that electrons collide elastically with atoms and molecules unless they have sufficient energy to raise the energy level of the atom to some higher state. Should a colliding electron possess more than the necessary energy, then it will rebound with the excess energy in kinetic form. It is evident that a critical potential measures the energy that an electron must possess to change the state of the atom, and it is convenient to express this energy in electron-volts. As 1 volt is 1/300 e.s.u. and the electronic charge is $4{\cdot}770 \times 10^{-10}$ e.s.u., the energy possessed by an electron which has experienced a free fall through a potential difference of 1 volt is $4{\cdot}770 \times 10^{-10}/300$ ergs or $1{\cdot}591 \times 10^{-12}$ ergs. Besides its application to critical potentials, the electron-volt, and its corresponding energy, is a unit widely employed in electronics and it is customary to speak of energies of so many "volts." With the foregoing definition in mind this should not cause any misunderstanding. Thus we may say that the energy required to ionize mercury vapour is 10·4 volts.

According to the quantum theory

$$hf = E_f - E_i = eV_c$$

If V_c is expressed in volts and, instead of f, the wave number is employed, we may write

$$V_c = 1{\cdot}2344 \times 10^{-4}\ \gamma$$

after substituting for h and e. As the wave number can be more accurately determined than the critical potential, it follows that once a spectral line has been definitely identified with a critical potential, the value of the latter may be more accurately found by calculation from the wave number of the spectral line.

In particular good results have been obtained with the alkali metals by the foregoing method. The vapours of these metals are monatomic and therefore tend to give simple spectra. An excitation potential has been found which corresponds to the first line in the principal spectral series, and also an ionization potential which agrees with that calculated from the wave number of the limit of the principal series. Table 1–1 permits a comparison of some calculated and observed results.

TABLE 1–1

ELEMENT	LIMIT WAVE NUMBER	V_i		FIRST LINE WAVE NUMBER	V_c	
		Cal.	Ob.		Cal.	Ob.
Sodium	41,449	5·116	5·13	16,973	2·095	2·12
Potassium	35,006	4·321	4·1	13,043	1·610	1·55
Rubidium	33,689	4·159	4·1	12,817	1·582	1·6
Caesium	31,405	3·877	3·9	11,732	1·448	1·48

It may be stated that observation has not yet been made of excitation potentials corresponding to all lines in the principal series, and, compared with the number of spectral lines observed, critical potentials are relatively few.

The Magneton

A consequence of Bohr's theory of the hydrogen atom is the magneton, the fundamental unit of magnetic moment. On page 141 it is shown that the motion of the electron in its orbit produces a magnetic moment equal to $ea/T = eav/2\pi r$ or $evr/2$. But from (1–76) $vr = nh/2\pi m$ and hence the moment is $(nh/4\pi)/e/m$. For $n = 1$

$$\mu = \frac{h}{4\pi} \cdot \frac{e}{m}$$

which is termed the *Bohr magneton*. Since the angular momenta of electrons are always multiples of $h/2\pi$, it follows that all magnetic moments are multiples of the magneton. Substituting for e/m and h we have $\mu = 9{\cdot}21 \times 10^{-21}$ e.m.u.

Elliptical Orbits

The circular orbits assumed for the hydrogen atom represent a special case, for the usual orbit under the inverse square law is an elliptical one. If an electron is moving in an elliptical orbit it has radial as well as angular momentum, because both the radius and angular position are varying. Hence the radial as well as the angular momentum must be quantized, and the phase integral applied to both degrees of freedom. Thus

$$J_\phi = \int p_\phi dq_\phi = n_\phi h$$

$$J_r = \int p_r dq_r = n_r h$$

where n_ϕ is the azimuthal and n_r the radial quantum number.

Investigation shows that these two numbers are not independent, but are related to the ellipse axes by the relation

$$\frac{b}{a} = \frac{n_\phi}{n_\phi + n_r}$$

where b and a are, respectively, the semi-minor and semi-major axes of the ellipse. Also the frequency for an electron transition between two orbits is found to be

$$f = \frac{2\pi^2 me^4}{h^3}\left[\frac{1}{(n_{\phi1} + n_{r1})^2} - \frac{1}{(n_{\phi2} + n_{r2})^2}\right]$$
$$= \frac{2\pi^2 me^4}{h^3}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

these equations being of the same form as that of (1–84).

Considering the forms of the ellipses involved, if in $n = (n_\phi + n_r)$, $n_r = 0$, $n = n_\phi$, and the orbit is circular. If $n_\phi = 0$ there is no azimuthal motion and the orbit is a straight line through the nucleus. Obviously this is inadmissible. Thus for a given value of n, there are n different shapes of ellipse determined by n_ϕ/n, commencing with $n_r = 0$, $n_\phi = n$, and passing through $n_r = 1$, $n_\phi = n - 1$, $n_r = 2$, $n_\phi = n - 2$, etc., up to $n_r = n - 1$ and $n_\phi = 1$. Subsequently the azimuthal quantum number was modified from n_ϕ to $l = (n_\phi - 1)$ where l takes the values 0, 1, 2, 3, 4, etc., when $n_\phi = 1, 2, 3, 4, 5$, etc.

Considering the case when $n = 1$, we have

$$n_\phi = 1,\ l = 0,\ n_r = 0$$

and the orbit is circular. This is the lower level for all K X-rays and for the Lyman series in the hydrogen atom.

When $n = 2$, we have

$$n_\phi = 2,\ l = 1,\ n_r = 0$$
$$n_\phi = 1,\ l = 0,\ n_r = 1$$

which shows there are two possible orbits, one circular and the other elliptical. These orbits are the lower level for all L X-rays and for the Balmer series in the hydrogen atom.

Continuing this process it will be found that for every principal quantum number, n, there are n possible electron orbits, one of which is circular and the others elliptical.

SPATIAL QUANTIZATION

Should there be a field of force in any given direction the orientation of the electronic orbits will be affected by this. Considering

Fig. 1–9, let *FF* mark the direction of the field, while α is the angle formed between the field direction and a projection p perpendicular to the plane of the orbit through its centre. As angular momentum is a vector quantity the projection p represents to a suitable scale the magnitude of this vector. The component of this vector along the field axis, i.e. the angular momentum p_ψ for the azimuthal motion projected on to the equatorial plane, is $p \cos \alpha = p_\psi$. As an electron in the orbit *ABCD* has three co-ordinates, these must be separately quantized as when dealing with the co-ordinates ϕ and r in the elliptical case. Hence we must have

$$\int_0^{2\pi} p_\psi d\psi = n_\psi h$$

and
$$p_\psi = \frac{n_\psi h}{2\pi}$$

where n_ψ is an integer sometimes termed the third or magnetic quantum number. Thus, as $p = n_\phi h/2\pi$, we have

$$\frac{n_\phi h}{2\pi} \cos \alpha = \frac{n_\psi h}{2\pi}$$

and
$$\frac{n_\psi}{n_\phi} = \cos \alpha$$

which gives the angles which p, and hence the orbit, can take up relative to the field direction *FF*. From this result it will be seen that α is restricted to relatively few values. If $n_\phi = 1$, $n_\psi = 0$, or ± 1 and the plane of the orbit either coincides with or is perpendicular to the direction *FF*. If $n_\phi = 2$, $n_\psi = 0, \pm 1$, or ± 2 and α can have the values $0, \pi, \pm \pi/3$, and $\pi/2$.

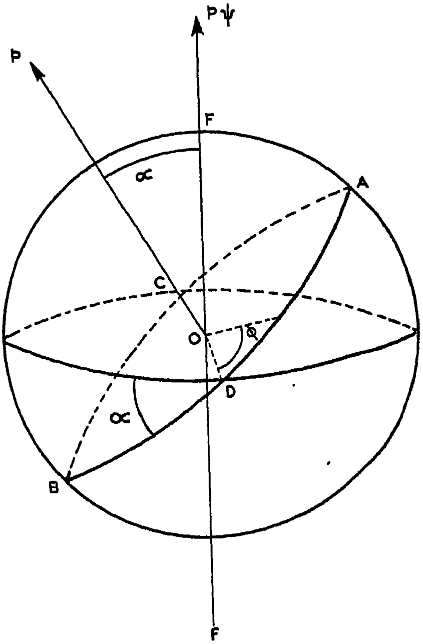


FIG. 1–9

ELECTRON SPIN

There is ample experimental evidence for spatial quantization, but such evidence gives somewhat different values for α than those already deduced. Thus, from spectroscopy it is found that for $n_\phi = 1$, $\cos \alpha = \pm 1$ instead of 0 and ± 1. For $n_\phi = 2$, $\cos \alpha = \pm \frac{1}{3}$ and ± 1, while for $n_\phi = 3$, $\cos \alpha = \pm 1/5$, $\pm 3/5$, and ± 1 instead of 0, $\pm \frac{1}{3}$, $\pm \frac{2}{3}$, and ± 1. If, instead of letting n_ϕ represent the orbital angular momentum, we use $j = l + \frac{1}{2}$, then agreement is found between the experimental and deduced values of $\cos \alpha$. For $j = l + \frac{1}{2}$, j will not have values 1, 2, 3, 4, etc., as with n_ϕ, but odd halves of integers such as $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, etc., when $l = 0, 1, 2$, etc. This means that the effective angular momentum of a single electron orbit is $jh/2\pi$ rather than $nh/2\pi$ with a least value of $h/4\pi$. Thus, the projected value of p on the field axis is $p_\psi = jh/2\pi \,.\, \cos \alpha$, and this leads to the values of α found by experiment.

In dealing with Bohr's theory of the hydrogen atom it was found that the various spectral series could be obtained by substituting different values for the principal quantum number, n. Close observation, however, reveals that each spectral line actually consists of two or three lines extremely close together. This feature is known as the *fine structure* of the spectral lines, two closely spaced lines being termed a doublet, and three, a triplet. This phenomenon is by no means confined to hydrogen, but is found with many other atoms. For example, the well-known D lines of sodium at 5890 Å and 5896 Å constitute a doublet. This multiplicity of lines, and hence energy values, is partly due to what is termed *electron spin.* This means that in order to account for the substitution of $j = l + \frac{1}{2}$ for n_ϕ and multiplets, it was assumed that the electron spins about its own axis. Hence the electron has an angular momentum about its axis and, since a spinning charge produces a field, the electron has also a magnetic moment. The magnetic moment is taken to be one Bohr magneton, i.e. $eh/4\pi m$, and the angular momentum one-half of that of an electron rotating on the innermost orbit of a hydrogen atom, i.e. $\frac{1}{2}(h/2\pi)$. As the electron may spin in either direction its spin number is written as $s = \pm \frac{1}{2}$. Hence the *total* angular momentum of the electron is given by $jh/2\pi = (l \pm s)\,(h/2\pi)$, this indicating the origin of the empirical quantity $l + \frac{1}{2}$ derived from spectroscopy.

One effect of electron spin is to produce a tendency for electrons to pair, two electrons with opposite spin numbers attempting to associate. This is, of course, due to the effects of the magnetic moments. It seems probable that a proton has also spin and magnetic moments. However, as μ varies as $1/m$, it appears that μ for the proton is no more than about 1/2000 of that for the electron, and, hence, may generally be neglected.

THE VECTOR MODEL ATOM

Considering a single electron orbit, the description of this must include three quantum numbers, viz. the principal quantum number n, the azimuthal quantum number l, and the spin number s. The orbital angular momentum is $p = lh/2\pi$, with l taking the values 0, 1, 2, 3, etc. This is a vector either upward or downward normal to the plane of the orbit as shown by Fig. 1–10. The spin vector is perpendicular to the plane of the orbit upward, as at (*a*), for $l + s$ and downward, as at (*b*), for $l - s$. The magnetic spin moment, which we may designate by μ_s, may either oppose the orbital magnetic moment or assist this, according to the direction of spin, for reference to Fig. 1–10 (*a*) and (*b*) shows these moments to be parallel. In general, the case shown at (*b*), i.e. $j = l - s$, is the more stable arrangement, the energy for (*b*) being less than for (*a*). Thus, for an orbit defined by n and Z the energy of the electron will have two values, one for $j = l + s$, and one for $j = l - s$. If the average energy of the electron in an orbit n is E, the doubling of j due to s changes E by amounts $+ dE_1$ and $- dE_2$. Hence transitions of the electron from the n-level to the ground state will produce pairs of spectral lines in the form of doublets.

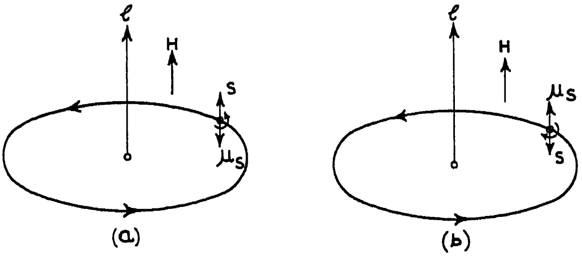


FIG. 1–10

ATOMIC STRUCTURE: ELECTRON ASSIGNMENT

As we have already seen, the atom is a kind of miniature planetary system with the nucleus situated at the focus of an elliptical orbit. In the special case of a circular orbit the nucleus is situated at the centre of the circle. In a multi-electron atom it is considered that the electrons revolve on a number of orbits of different radii, each collection of electrons with its particular orbit being known as a *shell*. Relative to a given field direction, the planes of the orbits will have various orientations and the angular momentum of the atom as a whole will be the vector sum of the angular momenta of the various orbits. The same consideration, of course, applies to the resultant spin and the magnetic moments.

Under spatial quantization it was shown that for each value of l (the azimuthal quantum number) a number of projections of $p = lh/2\pi$ on the field axis was possible. Instead of writing n_ψ as the magnetic quantum number, as we did when employing n_ϕ as the azimuthal quantum number, we shall now write m_l for this number. To find the number of possible values of m_l we may proceed as follows: As the angular momentum p is an integral value l of $h/2\pi$, it can have integral projections corresponding to $m_l = l$, $(l - 1)$, $(l - 2)$, $(l - 3)$, . . . 0 . . . , $-(l - 3)$, $-(l - 2)$, $-(l - 1)$, $-l$ times $h/2\pi$ on the field axis. This means that m_l can have $(2l + 1)$ values. For each value of l there are two values of j corresponding to the two electron spin moments. Hence when the *total* angular momentum of an orbit is projected on to the field axis there are $2(2l + 1)$ values of m_l.

In order to designate a possible electronic orbit or state in an atom, four numbers must be given—

1. The principal quantum number n.
2. The azimuthal quantum number l ($n = l$ possible values).
3. The magnetic quantum number m_l($2l + 1$ possible values).
4. The spin quantum number $m_s = \pm \frac{1}{2}$.

For a given value of n, the total number of different possible states is given by a series of n terms of the form $2(2l + 1)$ for each of the possible n values of l. Thus, the number of states is given by

$$2[2(0) + 1] + 2[2(1) + 1] + 2[2(2) + 1] + 2[2(3) + 1] \text{ etc. . . .}$$
$$= 2 + 6 + 10 + 14 + \text{etc. . .}$$

Thus, if $n = 1$, the number of states is $2 = 2(1)^2$, if $n = 2$, the number is $(2 + 6) = 8 = 2(2)^2$, if $n = 3$, the number is $(2 + 6 + 10) = 18 = 2(3)^2$, or for an orbit of principal quantum number n, the number of different possible states is $2n^2$.

PAULI'S EXCLUSION PRINCIPLE

According to the exclusion principle of Pauli no two electrons within the atom at the same instance can have the same values of n, l, m_l and m_s. From this restriction and the assumption that an electron will tend to occupy the position of lowest potential energy vacant, the manner in which the electrons are arranged within the atom can be found. Commencing with an atom for which $Z = 1$, i.e. hydrogen, and assuming it to be devoid of electrons, the addition of one electron produces a neutral hydrogen atom. The electron goes to the $n = 1$ state with $l = 0$ and $s = \pm \frac{1}{2}$. For He, $Z = 2$, and the first electron added has $n = 1$, $l = 0$, and $s = \pm \frac{1}{2}$. The He structure is a particularly stable one and is chemically inert, i.e. helium forms no compound with any other elements. The electrons in the $n = 1$ state form the K shell in all atoms, for when one of these electrons is raised to a higher energy level, its return gives rise to the K series X-rays, these appearing in all atoms beyond He.

In the case of Li, $Z = 3$. The first two electrons go to the $n = 1$ shell, which is then complete, because for $n = 1$ the number of possible states is equal to 2. The third electron must, therefore, commence a second shell outside the first. For this shell $n = 2$ and the electron orbit may either have $l = 0$ or $l = 1$. Spectroscopic data indicates that $l = 0$, which, as we have seen on page 48, gives an elliptical orbit.

For Be, $Z = 4$. The fourth electron has $n = 2$, $l = 0$, and a spin number of $s = - \frac{1}{2}$ if the third electron has $s = + \frac{1}{2}$, or vice versa.

For B, $Z = 5$. The first four electrons have the same states as with Be, except that their radii are smaller as Z is one greater. The fifth electron goes to the $n = 2$, $l = 1$ state and hence has a circular orbit.

For $l = 1$, there are six possible states; i.e. three circular orbits with the same radii but different orientations, each orbit being capable of holding two electrons with opposite spin numbers. Hence there are three values of the magnetic quantum number m_l, and two possible states to each value. From this it follows that the atoms C, $Z = 6$; N, $Z = 7$; O, $Z = 8$; F, $Z = 9$; and Ne, $Z = 10$; each add electrons of the type $n = 2$, $l = 1$, which, individually, take on one of the $2(2l + 1)$ states.

With Ne all the states corresponding to $n = 2$ (i.e. $2(2)^2$ states) are occupied by electrons, which means that we have two completed shells surrounding the nucleus. As in each orbit the electrons travel in opposite directions and have opposite spins, it follows that the resultant magnetic moment of Ne is zero. However, in an external field Ne shows diamagnetic behaviour due to the Larmor precession described in Chapter III. The shell corresponding to $n = 2$ is known as the L shell. That is, when an electron is removed from this shell its subsequent return produces the L series X-rays. Compared with the He atom the value of Z for Ne is five times as large. Hence the average displacement of the K shell from the nucleus in Ne is one-fifth of that in He.

Continuing the process of atom building, for Na, $Z = 11$ and the eleventh electron must have $n = 3$, i.e. the formation of a new shell, the M-shell, must be commenced. The value of l may be either 0, 1, or 2. Actually this electron has $l = 0$, $m_l = 0$ and $s = \pm \frac{1}{2}$, and hence circulates on one of the three elliptical orbits corresponding to $n = 3$.

The next atom is Mg with $Z = 12$ and for the twelfth electron we have $n = 3$, $l = 0$, $m_l = 0$ and a spin number of $s = -\frac{1}{2}$ if the eleventh electron has $s = \frac{1}{2}$, or $+\frac{1}{2}$ if vice versa. For Al, $Z = 13$ and the thirteenth electron has $n = 3$, $l = 1$. Now for $l = 1$ there are $2(2l + 1) = 6$ states and hence a further six electrons will be subsequently added as Z increases to eighteen for Ar. Now, although the M shell is not complete for Ar, nevertheless this element is particularly stable, forming, as He and Ne, no chemical compound with any other elements. That is, Ar is an inert gas.

Following Ar is K with $Z = 19$. For several reasons, mainly spectroscopic, it is considered that the nineteenth electron does not go to the $n = 3$, $l = 2$ state, but to the $n = 4$, $l = 0$ state. This means that this electron commences a new shell, the N shell. However, the characteristics of the elements Sc to Cu are such that it is considered that the $n = 3$, $l = 2$ state is recommenced with Sc and terminates with Cu. Thus from $Z = 21$ the nineteenth electron goes to the M shell. This process of starting a new shell before an old one is completed is not confined to the case just given, but occurs again with Rb, Cs, as shown by Table 1–2. It may be stated that the problem of why a new shell should be commenced before the old one is complete is not fully solved.

Table 1–2 gives a complete summary of the electron assignments of the ninety-two elements. Instead of writing $l = 0, 1, 2, 3$, etc., the following designation is used—

l	0	1	2	3
	s	p	d	f

Thus $3s$ means $n = 3$, $l = 0$, and, again, $2d$ indicates $n = 2$, $l = 2$. The number of electrons contained in a completed shell follows from Pauli's exclusion principle for the number of different states corresponding to a given value of n is $2n^2$. Thus, in completed K, L, M, and N shells there are respectively 2, 8, 18, and 32 electrons.

The reason there are no further elements beyond 92 is that

TABLE 1–2

		1s
1	H	1
2	He	2

		2s	2p
3	Li	1	
4	Be	2	
5	B	2	1
6	C	2	2
7	N	2	3
8	O	2	4
9	F	2	5
10	Ne	2	6

		3s	3p	3d	4s
11	Na	1			
12	Mg	2			
13	Al	2	1		
14	Si	2	2		
15	P	2	3		
16	S	2	4		
17	Cl	2	5		
18	A	2	6		
19	K	2	6		1
20	Ca	2	6		2
21	Sc	2	6	1	2
22	Ti	2	6	2	2
23	V	2	6	3	2
24	Cr	2	6	4	2
25	Mn	2	6	5	2
26	Fe	2	6	6	2
27	Co	2	6	7	2
28	Ni	2	6	8	2
29	Cu	2	6	10	1

		4s	4p	4d	5s
30	Zn	2			
31	Ga	2	1		
32	Ge	2	2		
33	As	2	3		
34	Se	2	4		
35	Br	2	5		
36	Kr	2	6		
37	Rb	2	6		1
38	Sr	2	6		2
39	Y	2	6	1	2
40	Zr	2	6	2	2
41	Nb	2	6	4	1
42	Mo	2	6	5	1
43	Ma	2	6	(5)	(2)
44	Ru	2	6	7	1
45	Rh	2	6	8	1
46	Pd	2	6	10	
47	Ag	2	6	10	1
48	Cd	2	6	10	2

		4s	4p	4d	4f	5s	5p	5d	6s
49	In	2	6	10		2	1		
50	Sn	2	6	10		2	2		
51	Sb	2	6	10		2	3		
52	Te	2	6	10		2	4		
53	I	2	6	10		2	5		
54	X	2	6	10		2	6		
55	Cs	2	6	10		2	6		1
56	Ba	2	6	10		2	6		2
57	La	2	6	10		2	6	1	2
58	Ce	2	6	10	(1)	2	6	(2)	(1)
59	Pr	2	6	10	(2)	2	6	(2)	(1)
60	Nd	2	6	10	(3)	2	6	(2)	(1)
61	Il	2	6	10	(4)	2	6	(2)	(1)
62	Sm	2	6	10	(5)	2	6	(2)	(1)
63	Eu	2	6	10	(6)	2	6	(2)	(1)
64	Gd	2	6	10	(7)	2	6	(2)	(1)
65	Tb	2	6	10	(8)	2	6	(2)	(1)
66	Dy	2	6	10	(9)	2	6	(2)	(1)
67	Ho	2	6	10	(10)	2	6	(2)	(1)
68	Er	2	6	10	(11)	2	6	(2)	(1)
69	Tm	2	6	10	(12)	2	6	(2)	(1)
70	Yb	2	6	10	(13)	2	6	(2)	(1)
71	Lu	2	6	10	(14)	2	6	(2)	(1)

		5s	5p	5d	6s	6p	6d	7s
72	Hf	2	6	(2)	(2)			
73	Ta	2	6	(3)	(2)			
74	W	2	6	4	2			
75	Re	2	6	(5)	(2)			
76	Os	2	6	(6)	(2)			
77	Ir	2	6	(7)	(2)			
78	Pt	2	6	8	2			
79	Au	2	6	10	1			
80	Hg	2	6	10	2			
81	Tl	2	6	10	2	1		
82	Pb	2	6	10	2	2		
83	Bi	2	6	10	2	3		
84	Po	2	6	10	2	4		
85	—	2	6	10	2	5		
86	Rn	2	6	10	2	6		
87	—	2	6	10	2	6		1
88	Ra	2	6	10	2	6		2
89	Ac	2	6	10	2	6	(1)	2
90	Th	2	6	10	2	6	(2)	(2)
91	Pa	2	6	10	2	6	(3)	(2)
92	U	2	6	10	2	6	(5)	(1)

Closed shells are not repeated in the succeeding part of the table. Numbers in parentheses are conjectural.

from 84 onwards the elements are radio-active. Hence, presumably, anything beyond an atomic number of 92 is too unstable to exist.

General References

A Treatise on Algebra, C. Smith (Macmillan).
Integral Calculus, B. Williamson (Longmans, Green).
Kinetic Theory of Gases, E. Bloch (Methuen).
Theory of Heat, T. Preston (Macmillan).
Heat, J. H. Poynting and J. J. Thomson (Griffin).
Atomic Structure, L. B. Loeb (Chapman & Hall).

CHAPTER II

GASEOUS ELECTRICAL CONDUCTION

UNDER normal conditions and relatively low potential gradients (20 to 30 volts per cm.) gases are relatively poor conductors. In fact, for the majority of purposes they may be regarded as perfect insulators. However, if a pair of electrodes is placed in a gas and a potential gradient produced in the latter, a sufficiently sensitive instrument will record the presence of a current. As the potential is raised from zero the current initially increases but shortly reaches a state of saturation. In order to demonstrate and measure this current the apparatus of Fig. 2–1 may be employed. It consists of a gold-leaf electroscope fitted to the interior of a glass vessel, the inner surface of which is silvered to make it conducting. *S* is a sulphur block and *W* an iron wire. On attracting *W* with a magnet it makes contact with the metal support of the gold leaves and a potential difference equal to that of the battery is set up between the leaves and the interior of the vessel. Due to the conductivity of the gas the charge on the leaves leaks away.

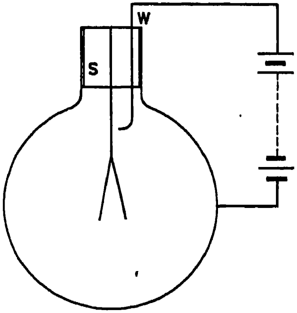


FIG. 2–1

Now
$$i = \frac{dQ}{dt} = C\frac{dV}{dt}$$

or
$$i_{av} = C\frac{V}{t}$$

where i_{av} is the average value of the current, C the capacity of the apparatus, and V the change in potential in a time t. The change in V may be found by plotting the position of the leaves against known potentials. The value of i is found to be about $10^{-8}v$ e.s.u., where v is the volume of the gas. If we have 1 litre of gas this gives 10^{-5} e.s.u. of current, or $10^{-5}/3 \times 10^{10}$ e.m.u. $= \frac{1}{3}\, 10^{-15}$ e.m.u., or $3{\cdot}3 \times 10^{-15}$ amp. This clearly demonstrates the extremely low conductivity of a gas under normal conditions.

Current Measurements

The foregoing experiment may serve to show the extremely small currents which may be encountered in gaseous conduction, and it is desirable before proceeding to indicate methods of measuring such currents. In many cases some form of galvanometer may be employed and the suitability of this method or otherwise may be judged by what is known as the "figure of merit" of the galvanometer. This figure refers to the current necessary to deflect a reflecting instrument through 1 mm. on a scale placed at a distance of 1 metre from the instrument. Some typical figures are shown by Table 2–1.

TABLE 2–1

Type	Figure of Merit in Amperes
Ayrton-Mather Moving Coil .	10^{-6} to 10^{-9}
Thomson	10^{-9} to 10^{-11}
Broca	10^{-10}
Einthoven	10^{-12}
Thermo-galvanometer . .	2×10^{-6}

It is apparent from the above table that none of these instruments is sufficiently sensitive to measure the current in the experimental arrangement of Fig. 2–1. In such cases an electrometer or tilted electroscope must be used. The former has a capacity of about 50 e.s.u. and a sensitivity of 1000 mm. per volt, and the latter a capacity of one or two e.s.u. and a sensitivity of 200 mm. per volt. Assuming that the electrometer has a sensitivity of 1000 divisions

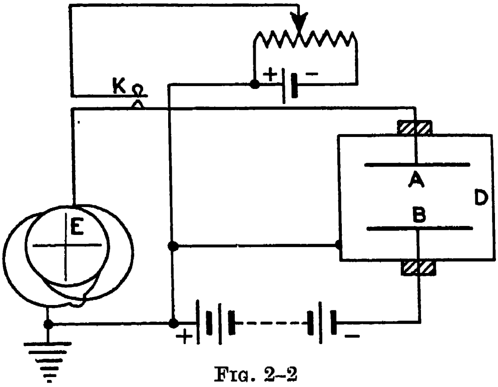


Fig. 2–2

per volt and that a rate of deflexion of 1 division per 10 sec. can be observed, then $dV/dt = 10^{-4}$. If the capacity of the instrument is 50 e.s.u., this is $50/(9 \times 10^{11}) = 5{\cdot}55 \times 10^{-11}$ farads. Therefore the smallest current which can be observed is $5{\cdot}55 \times 10^{-11} \times 10^{-4} = 5{\cdot}5 \times 10^{-15}$ amp. Similarly for the electroscope $dV/dt = 0{\cdot}5 \times 10^{-3}$. If the capacity is 2 e.s.u., this is $2{\cdot}22 \times 10^{-12}$ farad. Therefore the smallest observable current is $0{\cdot}5 \times 10^{-3} \times 2{\cdot}22 \times 10^{-12} = 1{\cdot}11 \times 10^{-15}$ amp.

In order to demonstrate the application of either of the above instruments, say the electrometer, the arrangement of Fig. 2–2 may be considered. Within the chamber D are two parallel-plane electrodes one of which, B, differs in potential by some few hundred volts from A. When K is depressed there is no potential across the electrometer quadrants E. On releasing K, A starts to acquire ions of the same sign as the potential of B, and thus the potential difference between A and B commences to change, this change being recorded by E. If t is the time for a deflexion V then

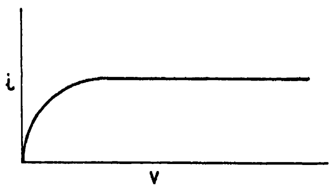


FIG. 2–3

$$i_{av} = C\frac{V}{t}$$

and, providing V is small, this current corresponds to the potential difference between A and B. By varying the potential of B the magnitude of the current between the planes can be studied for various potential differences between A and B. A curve of the form shown by Fig. 2–3 is obtained.*

Theory of Conduction Through Gases: The Separately Maintained Discharge

The existence of the current in the foregoing experiment is, of course, due to the presence of ions and electrons within the gas. These are due to the continual presence of various ionizing agents, such as traces of radio-active substances, cosmic radiation, ultra-violet light, etc. Let it be assumed that the ionizing agent is producing, uniformly, p pairs of ions per c.c. per sec. in the gas. Now ions of opposite sign will tend to recombine. The time that a given positive ion will take to recombine will be inversely proportional to the number of negative ions present per cubic centimetre

*The potentiometer in Fig. 2–2 is for calibrating.

of gas, while the numbers of positive ions recombining in a given time is proportional to the number present. Hence the rate of recombination is proportional to n^2 where n is the number of ions of either polarity present per cubic centimetre of gas. It may be written as αn^2 where α is termed the coefficient of recombination. In the absence of a field within the gas the rate of increase of the number of ions per cubic centimetre is equal to the number formed per second minus the number recombining per second, or

$$\frac{dn}{dt} = p - \alpha n^2$$

When a steady state is reached

$$\frac{dn}{dt} = 0 \text{ and } p = \alpha n^2$$

If the source of ionization is removed

$$p = 0 \text{ and } \frac{dn}{dt} = - \alpha n^2$$

$$\frac{dn}{n^2} = - \alpha dt$$

$$\int \frac{dn}{n^2} = - \alpha \int dt$$

$$-\frac{1}{n} + K = - \alpha t$$

When $t = 0$, $K = 1/n_0$, where n_0 is the value of the ionization when the source is removed. Hence we have

$$n = \frac{n_0}{1 + \alpha n_0 t}$$

MOTION OF IONS UNDER A UNIFORM ELECTRIC FIELD

If X denotes the electric field strength, then the force on a charged particle is Xq, where q is the particle's charge. If m is the mass of the particle then the initial acceleration imparted to it by the field is Xq/m and if its passage were unimpeded, it would acquire a velocity in a time t equal to

$$v = \frac{Xqt}{m} \quad . \quad . \quad . \quad . \quad (2\text{–}1)$$

Actually, however, the ion cannot travel far before colliding with other molecules and in general it may be assumed that the velocity component produced by the field is destroyed after the ion has travelled a distance equal to the mean free path of the gas. Now, from page 18 the time between collisions is $l/\overline{C}$ and hence (2–1) may be written

$$v = \frac{Xql}{m\overline{C}}$$

and the average velocity added by the field is

$$v_{av} = \frac{Xql}{2m\overline{C}} \quad . \quad . \quad . \quad . \quad (2\text{–}2)$$

From (1–34) we have

$$\tfrac{1}{2}mC^2 = \frac{3}{2}kT$$

or

$$\frac{3\pi m\overline{C}}{16} = \frac{3}{2}kT$$

and

$$\overline{C} = \sqrt{\frac{8kT}{m\pi}}$$

Substituting in (2–2)

$$v_{av} = \frac{1}{4}\sqrt{\frac{\pi}{2}}\frac{Xql}{\sqrt{mkT}}$$

If u and v are, respectively, the actual velocities of the positive and negative ions in a field of strength X, then

$$u = K_1X, \; v = K_2X$$

where K_1 and K_2 are constants for a given gas under given conditions of temperature and pressure. When the field is measured in volts per centimetre they are known as the mobilities of the ions.

CURRENT THROUGH A GAS

Let a potential difference of V be applied between a pair of parallel-plane electrodes within a gas. Considering a plane in the gas normal to the field, in one second all the positive ions situated at a distance of less than u cm. from it will be driven across the plane, and all the negative ions less than v cm. distant will be driven across it in the other direction. If the charge q is the same for all the ions the total transference across the plane is $nq(u + v)$ units per sq. cm. per sec., or if A is the area of the plane,

$$i = Anq(u + v) = AnqX(K_1 + K_2)$$

The determination of the current i is difficult because the presence of a current implies the withdrawal of ions from the gas. However, i may be readily determined in certain limiting cases.

(1) The current is so small that the number of ions withdrawn is negligible compared with the number per cubic centimetre. In this case the loss of ions is solely due to recombination. As the electrodes are assumed to be parallel-planes, the field strength $X = V/d$, where d is the distance between the electrodes. Hence

$$= \frac{AnqV}{d} (K_1 + K_2)$$

and if n is constant the current obeys Ohm's Law.

(2) The strength of the field is so large that the ions are withdrawn to the electrodes without appreciable recombination occurring. If B is the volume of gas between the electrodes, the number of ions is Bp and the total charge conveyed to either electrode in one second is Bpq. This is the maximum current through the gas, and, provided the voltage is sufficient to produce saturation, is independent of the voltage between the electrodes.

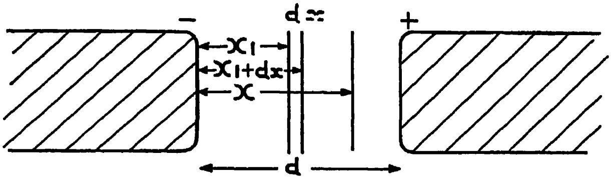


Fig. 2–4

Cases (1) and (2) above, of course, explain the curve of Fig. 2–3. The initial portion of this corresponds to (1) (Ohm's Law) and the portion parallel to the abscissa to (2).

The type of discharge so far considered is termed "separately maintained" because it depends on an external source of ionization. A further form of this type of discharge will now be considered in which ionization by collision occurs.

If after saturation has been reached the potential between the electrodes is continually increased, it is found that the curve of Fig. 2–3 turns sharply upwards, denoting a rapid increase in the current through the gas. This may be explained on the supposition that an ion in traversing the mean free path of the gas has also fallen through a voltage at least equal to the ionization potential of the gas. Should this be so, then ionization by collision will occur with the production of fresh pairs of ions. Initially, at least, this process will be due to electrons because of their very much higher mobility than positive or negative ions.

Considering Fig. 2–4, let n_0 be the number of ions produced by the ionizing agent per cubic centimetre of gas per second at a distance x_1 from the negative electrode. If α is the number of pairs of ions and electrons produced by each electron per centimetre length of its path, then the number of fresh ions of either sign produced in a distance dx is $\alpha n_- dx$. Hence the rate at which electrons are increasing as the positive electrode is approached is

$$\frac{dn_-}{dx} = \alpha n_-$$

and the number of electrons passing per second through a plane at a distance x from the cathode and beyond x_1 is given by

$$\int \frac{dn_-}{n_-} = \alpha \int_{x_1}^{x} dx$$

$$\log \frac{n_-}{n_0} = \alpha(x - x_1)$$

and

$$n_- = n_0 \varepsilon^{\alpha(x - x_1)} \quad . \quad . \quad . \quad (2\text{–}3)$$

The number of positive ions is found from

$$\begin{aligned} dn_+ &= -\alpha n_- dx \\ &= -\alpha n_0 \varepsilon^{\alpha(x - x_1)} dx \\ n_+ &= -n_0 \varepsilon^{\alpha(x - x_1)} + K \end{aligned}$$

When $x = d$, $n_+ = 0$ and thus

$$n_+ = n_0[\varepsilon^{\alpha(d - x_1)} - \varepsilon^{\alpha(x - x_1)}] \quad . \quad . \quad (2\text{–}4)$$

Multiplying (2–3) and (2–4) by the electronic charge e, these equations may be transformed into those for current, and we have

$$i_- = i_0 \varepsilon^{\alpha(x - x_1)} \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}5)$$

$$i_+ = i_0[\varepsilon^{\alpha(d - x_1)} - \varepsilon^{\alpha(x - x_1)}] \quad . \quad . \quad . \quad (2\text{–}6)$$

The total current is the sum of (2–5) and (2–6) and thus

$$I = i_- + i_+ = i_0 \varepsilon^{\alpha(d - x_1)} \quad . \quad . \quad . \quad (2\text{–}7)$$

Considering now the case where the gas is uniformly ionized throughout its volume, the number of electrons formed by the ionizing agency between two planes x and $x + dx$ from the negative electrode is $n_0 dx$. The number of electrons reaching the positive electrode due to this layer is $n_0\varepsilon^{\alpha(d-x)}\,dx$ and hence the total number of electrons arriving at the positive electrode per second is

$$n_0\int_0^d \varepsilon^{\alpha(d-x)}\,dx = \frac{n_0}{\alpha}(\varepsilon^{\alpha d} - 1)$$

or

$$i = \frac{i_0}{\alpha}(\varepsilon^{\alpha d} - 1)$$

An arrangement of some importance is where the ionization is produced at the cathode, say by the action of ultra-violet radiation on this. We then have $x_1 = 0$ and

$$n_- = n_0\varepsilon^{\alpha x} \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}8)$$

$$n_+ = n_0[\varepsilon^{\alpha d} - \varepsilon^{\alpha x}] \quad . \quad . \quad . \quad (2\text{–}9)$$

$$i_- = i_0\varepsilon^{\alpha x} \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}10)$$

$$i_+ = i_0[\varepsilon^{\alpha d} - \varepsilon^{\alpha x}] \quad . \quad . \quad . \quad (2\text{–}11)$$

$$I = i_0\varepsilon^{\alpha d} \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}12)$$

where n_0 is the number of electrons produced per second at the cathode and i_0 is the corresponding current.

The foregoing expressions ignore recombination, it being assumed that the ions and electrons are swept to the electrodes as rapidly as they are formed.

Breakdown: The Self-maintained Discharge

The formulae just deduced (which have been substantiated by experiment) indicate a current which depends upon and is proportional to the effect of the ionizing agent. If the voltage across a gas is sufficiently increased, a stage is reached at which the discharge no longer depends on i_0, which can then, if desired, be terminated without terminating the discharge. The discharge is then termed self-maintained, well-known examples being the spark and glow. It will now be shown that if positive ions can produce further ions and electrons, given certain conditions, a self-maintained discharge may result.

Let it be assumed that each positive ion produces β ions and electrons per centimetre length of its path. Then, referring to Fig. 2–4,

$$dn_- = \alpha(n_0 + n)dx + \beta n_+ dx \quad . \quad . \quad (2\text{–}13)$$

The number of electrons arriving at the positive electrode per second is

$$n = n_0 + n_- + n_+ \quad . \quad . \quad . \quad (2\text{–}14)$$

where n_0 is the number of electrons produced at the cathode per second by the ionizing agent. Substituting in (2–13) for n_+

$$dn_- = \alpha(n_0 + n_-)dx + \beta(n - n_0 - n_-)dx$$

and
$$\frac{dn_-}{dx} = (\alpha - \beta)(n_0 + n_-) + \beta n$$

from which
$$n_- = -n_0 - \frac{\beta}{\alpha - \beta} n + A\varepsilon^{(\alpha - \beta)x}$$

where A is a constant.

When $x = 0$, $n_- = 0$, therefore

$$A = n_0 + \frac{\beta}{\alpha - \beta} n$$

When $x = d$, $n_+ = 0$, therefore

$$n = \left(n_0 + \frac{\beta}{\alpha - \beta} n\right)[\varepsilon^{(\alpha - \beta)d} - 1] + n_0$$

or
$$n = n_0 \frac{(\alpha - \beta)\varepsilon^{(\alpha - \beta)d}}{\alpha - \beta\varepsilon^{(\alpha - \beta)d}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}15)$$

Multiplying both sides by the electronic charge e we have

$$I = i_0 \frac{(\alpha - \beta)\varepsilon^{(\alpha - \beta)d}}{\alpha - \beta\varepsilon^{(\alpha - \beta)d}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}16)$$

It is evident from this equation that if

$$\beta\varepsilon^{(\alpha - \beta)d} = \alpha \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}17)$$

a mathematically infinite current will result or a finite current can exist even if i_0 be made zero. This condition is known as breakdown of the gas. In effect this means that the positive ions are responsible for an internal source of electrons necessary for the self-maintenance of the discharge. However, there is not unanimous agreement regarding the manner in which these electrons are produced. The method just assumed, i.e. positive ion bombardment of gas molecules, has been objected to because of the relatively low velocities acquired by the positive ions. A possibility is the liberation of electrons from the cathode (secondary emission) by positive ion bombardment of this. From (2–11) the positive ion current at the cathode is

$$i_+ = i_0(\varepsilon^{\alpha d} - 1)$$

Suppose now that i_0 is the total current at the cathode and i' the current produced by the external agent. Then

$$i_0 = i' + \gamma i_0(\varepsilon^{\alpha d} - 1) \quad . \quad . \quad . \quad (2\text{–}18)$$

where γ is a constant of secondary emission. Substituting in (2–12) we have

$$I = i' \frac{\varepsilon^{\alpha d}}{1 - \gamma(\varepsilon^{\alpha d} - 1)} \quad . \quad . \quad . \quad (2\text{–}19)$$

Now if

$$\gamma(\varepsilon^{\alpha d} - 1) = 1 \quad . \quad . \quad . \quad (2\text{–}20)$$

a mathematically infinite current again results, the conditions being similar to those expressed by (2–16). It will be noted that (2–19) and (2–16) are identical if in the latter α is substituted for $(\alpha - \beta)$ and γ for $\beta/(\alpha - \beta)$. Hence without further information regarding β and γ the two processes of breakdown are indistinguishable.

For the last described process, self-maintenance necessitates that the positive ions which are generated by an electron in moving from cathode to anode regenerate at least one electron by means of secondary emission at the cathode. This necessity is implicit in (2–19), as will now be shown. Referring to (2–9), if n_0 electrons leave the cathode per second, then these generate $n_0(\varepsilon^{\alpha d} - 1)$ positive ions per second in their passage to the anode. Hence each electron produces $(\varepsilon^{\alpha d} - 1)$ positive ions in the body of the gas and these in turn eject $\gamma(\varepsilon^{\alpha d} - 1)$ electrons from the cathode. Now if this last quantity equals unity, then an electron initially emitted by the cathode has liberated another electron from the same site. This in turn will produce a third electron, the process continuing indefinitely with no necessity for an external ionizing agent. The criterion for this is

$$\gamma(\varepsilon^{\alpha d} - 1) = 1$$

as previously indicated by (2–19).

In addition to the two foregoing explanations of breakdown, other possibilities which may be here noted are the emission of electrons from the cathode and gas by photo-excitation by excited gas molecules and the influence of metastable atoms in the vicinity of the cathode.

Value of α and β: Paschen's Law

The value of α has previously been defined as the number of pairs of electrons and ions produced by an electron per centimetre length of its path. Evidently α will depend on the collision frequency of the electron and the energy it possesses on encountering a gas molecule. The latter quantity is

$$\tfrac{1}{2}mv^2 = \tfrac{1}{2}m\left(\frac{lXe}{m\bar{C}}\right)^2.$$

where l is the mean free path of the electron within the gas. From (1–45) l is inversely proportional to the gas density and pressure, and thus the electronic energy is proportional to X^2/p^2. From (1–46) the collision frequency is proportional to the gas density and pressure and thus α is proportional to p and some function of X/p. Hence we may write

$$\frac{\alpha}{p} = f_1\left(\frac{X}{p}\right) \quad . \quad . \quad . \quad . \quad (2\text{–}21)$$

Similarly for positive ions

$$\frac{\beta}{p} = f_2\left(\frac{X}{p}\right) \quad . \quad . \quad . \quad . \quad (2\text{–}22)$$

A probable form of the function f_1 may be derived as follows. The number of electrons generated by ionization within a lamina dx is $\alpha n dx$. This number may be considered proportional to n_1, the number arriving at x with energy greater than the ionizing potential, and to p the gas pressure. Therefore

$$\alpha n dx \propto p n_1$$

or

$$\alpha = ap\,\frac{n_1}{n}$$

where a is a constant. Now let N_0 electrons start from a plane with zero velocity, N arriving at a plane x. Those which make inelastic collisions in the next lamina dx may be written

$$pf(V)Ndx \quad . \quad . \quad . \quad . \quad (2\text{–}23)$$

where $V = Xx$ and $f(V)$ is the probability of an inelastic collision. As $dV = Xdx$, (2–23) may be written

$$dN = -\frac{p}{X}\,Nf(V)dV \quad . \quad . \quad . \quad (2\text{–}24)$$

the negative sign indicating a decrease in N with an increase in x and V. Integrating (2–24) and remembering that when x and $V = 0$, $N = N_0$

$$N = N_0\varepsilon^{-\frac{p}{X}\int_0^V f(V)dV} \quad .$$

Substituting N/N_0 for n_1/n and writing $\int_0^V f(V)dV = f_i(V_i)$ where V_i is the ionizing potential,

$$\alpha = Kp\varepsilon^{-(p/X)f_i(V_i)}$$

or

$$\frac{\alpha}{p} = Ae^{-Bp/X} \quad . \quad . \quad . \quad . \quad (2\text{–}25)$$

where A and B are constants for a given gas.

In order to test the validity of (2–25), it is necessary to employ a tube with either parallel-plane electrodes, or spheres the radius of which is large compared with the spacing between them. In these circumstances α is independent of x. A plot of α/p against p/X is shown in Fig. 2–5.

It was discovered empirically by Paschen that, for electrodes of the form just mentioned, the breakdown potential of a gas is a function of the gas pressure and the electrode spacing only, i.e. it is a function of pd. This may also be deduced theoretically in the following manner. From (2–21) and (2–22)

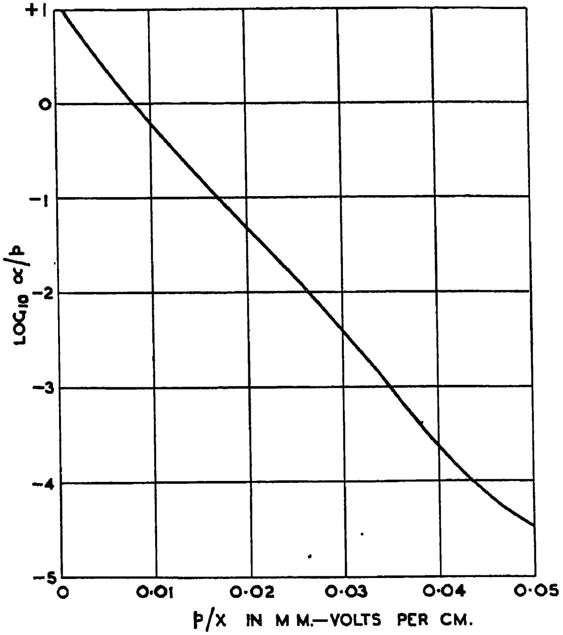


FIG. 2–5

$$\alpha = pf_1\left(\frac{V}{pd}\right) \quad . \quad . \quad . \quad . \quad (2\text{–}26)$$

$$\beta = pf_2\left(\frac{V}{pd}\right) \quad . \quad . \quad . \quad . \quad (2\text{–}27)$$

where instead of X we have written V/d, where V is the potential difference between the electrodes. From (2–17) the condition of breakdown is

$$\frac{\alpha}{\beta} = \varepsilon^{(\alpha - \beta)d}$$

or, from (2–26) and (2–27),

$$\frac{f_1\left(\dfrac{V}{pd}\right)}{f_2\left(\dfrac{V}{pd}\right)} = \varepsilon^{pd[f_1(V/pd) - f_2(V/pd)]}$$

from which

$$pd = \frac{\log \dfrac{f_1\left(\dfrac{V}{pd}\right)}{f_2\left(\dfrac{V}{pd}\right)}}{f_1(V/pd) - f_2(V/pd)}$$

showing that V is a function of pd only. The same law may also be deduced for the case expressed by (2–19) on the assumption that $\gamma = f_2(X/p)$.

Breakdown in Uniform Fields

An experimentally determined curve for the breakdown of air in a uniform field is shown by Fig. 2–6. A feature of great interest and importance in this is the minimum exhibited at about 350 volts. If the curve is obtained with spheres of large radius, any attempt at reducing pd below the minimum value is followed by breakdown occurring elsewhere than at the point of minimum electrode separation. This, of course, means that while V is held at what is known as the minimum sparking potential it is impossible to obtain breakdown at a value of pd smaller than indicated by Fig. 2–6. Conversely, it is impossible to produce breakdown in air under any circumstances at a voltage less than 350 volts. A further point of importance in Fig. 2–6 is the breakdown potential at atmospheric pressure. It will be seen that this is approximately 30,000 volts per cm.

The cause of the minimum sparking potential may be explained in the following manner. Assuming that d is fixed and p varied, at very low pressures the paucity of molecules is such that insufficient exist to produce adequate secondary emission unless the energy of the electrons is high. This, of course, necessitates a large value of

X, and consequently of V, and explains the left-hand branch of the curve of Fig. 2–6. At high values of p the mean free path of the electrons is small and insufficient energy for ionization is acquired between collisions unless X and V are again high. Between these two extremes it is evident that some minimum potential will exist, causing breakdown.

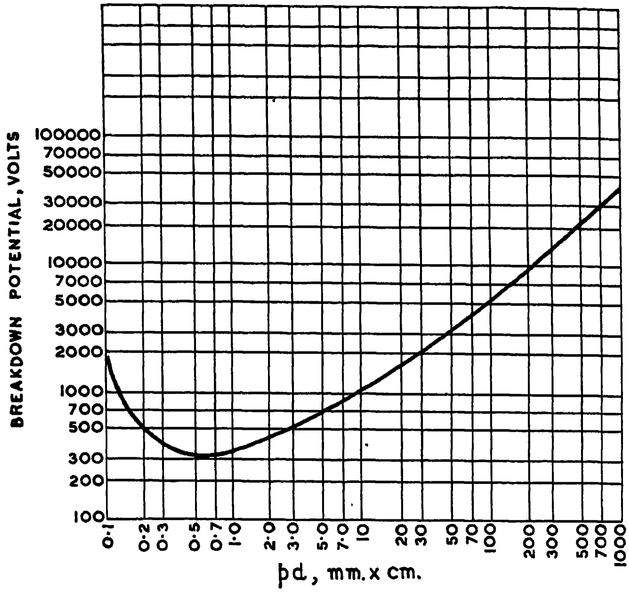


Fig. 2–6

Non-uniform Fields

The instances of gaseous conduction so far considered are the relatively simple ones involving a uniform field. Where the field is non-uniform (practically all cases except those employing parallel-plane electrodes or spheres whose r/d is large) conditions are much less simple and mathematical treatment cannot be so readily employed. In such cases α and β depend on x and in general we must write

$$\alpha = f_1(x)$$
$$\beta = f_2(x)$$

Thus for a non-uniform field, (2–12), should be written

$$I = i_0 \varepsilon^{\int_0^d f_1(x)dx}$$

In the case of breakdown due to liberation of electrons at the cathode by secondary emission, (2–20) must be written

$$\gamma(\varepsilon^{\int_0^d f_1(x)dx} - 1) = 1 \quad . \quad . \quad . \quad (2\text{–}28)$$

and provided the integration can be performed the breakdown condition may be determined. However, in the majority of cases this cannot be effected and thus we are limited to general considerations largely based on experimental results.

Principle of Similarity

Paschen's Law may be extended to cover the case of non-uniform fields, when it is known as the Principle of Similarity. This states that the breakdown voltage is a function of the gas density and linear dimensions for geometrically similar systems. Thus, if all the linear dimensions of a discharge system are multiplied by a factor K and the gas pressure is multiplied by $1/K$, the breakdown voltage remains unchanged. Consider an element of the discharge path dl at the extremities of which the potentials are v and $v + dv$. Then after multiplying the system by K the potential difference dv will exist over a path Kdl if the same potential difference is maintained between the electrodes. Hence the average field strength over this path is $1/K$ of that over the path dl. If the gas pressure is altered to $1/K$ of its former value, then the mean free path of the electrons will be K times as great and the energy per mean free path will be the same in the changed system as before. But the number of mean free paths per centimetre will be $1/K$ of the former value and hence the ionizing collisions over the paths dl and Kdl are, respectively,

$$\alpha dl \text{ and } \alpha K dl/K$$

or the ionizing collisions under the same potential difference dv are the same. Hence both systems will break down under the same potential between the electrodes.

In cases where the electrodes are of different dimensions (such as a point and plane) the field strength is greatest at the smaller. Hence ionization will be most intense here and breakdown will commence at this electrode. At points remote from the electrodes the field is always less than at these, and hence breakdown always commences close to the electrode surfaces. It may be noted that it is not necessary for ionization to occur over the whole discharge path for breakdown to take place. Provided breakdown conditions are realized where ionization does occur, then a breakdown will follow. This means that if the total discharge path is d_2, breakdown will occur if (2–17) is satisfied over a distance $x = 0$ to $x = d_1$ where $d_1 < d_2$. Hence that part of the path $d_2 - d_1$ will carry a current although ionization does not there occur.

It is shown by Fig. 2–6 that provided the electrode area is large compared with the spacing, the breakdown potential for air is approximately 30 kV per cm. From this it is concluded that in any discharge system, if the voltage gradient does not equal or exceed this value, breakdown will not occur at atmospheric pressure. Further, it is assumed that when breakdown takes place in a non-uniform discharge system ionization occurs between that electrode and the point at which the gradient falls to 30 kV per cm.

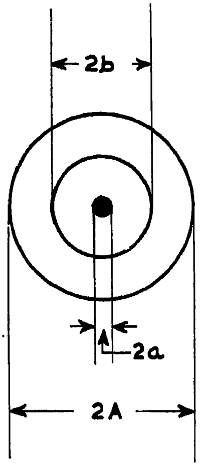


Fig. 2–7

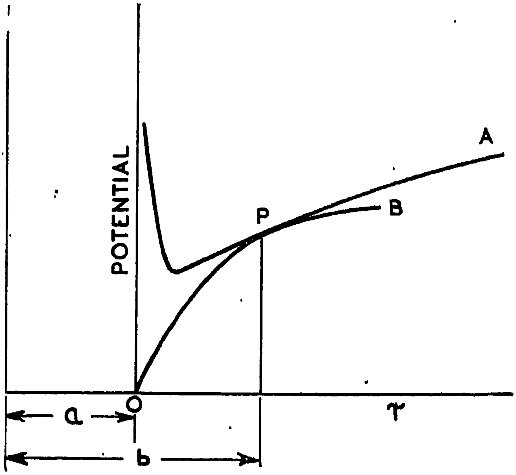


Fig. 2–8

The Corona Discharge

When breakdown occurs in uniform fields at atmospheric pressure, it usually takes the form of the well-known spark discharge. However, should the electrodes be small compared with their spacing, a local discharge may occur near their surfaces, this being termed a corona or brush discharge. Such discharges are characterized by faint luminosity near the electrodes, the production in air of ozone, a stream of air often issuing from the electrodes, and slight noise. The air in the immediate vicinity of the electrodes is, no doubt, highly conducting and increases the electrode area. This is sometimes referred to as the coronal extension of a conductor. The increased area also reduces the electrode spacing, a result being an increased uniformity in the field distribution. This process may continue to the point that the uniformity is such that breakdown in the form of a spark results.

To further consideration of the above subject the case of an electrode system consisting of a centrally placed wire within a cylinder will be investigated. Referring to Fig. 2–7, if X_s is the field strength at the surface of the wire, the field at any radius r is

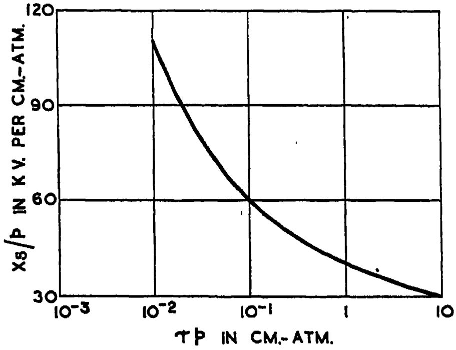


Fig. 2–9

$$X = X_s \frac{a}{r}$$

The potential V corresponding to X is

$$V = \int_a^r X dr = aX_s \log \frac{r}{a} \quad . \quad . \quad . \quad (2\text{–}29)$$

where V is the potential at any radius r. We shall now assume that ionization is confined within a radius b and that the potential difference across $(b - a)$ is the same as that needed to cause breakdown across the same length in a uniform field. Referring to Fig. 2–8, Curve A is the normal breakdown curve for uniform fields, while Curve B is plotted from (2–29), X_s being so chosen that the two curves have a single point of contact as shown. Hence P gives the distance $(b - a)$ and the construction of Fig. 2–8 permits the determination of X_s, i.e. the breakdown gradient at the surface of the wire. Fig. 2–9 shows breakdown potential plotted against wire diameter and indicates the very high values which may occur with non-uniform fields compared with those of an opposite character. Townsend gives the breakdown surface gradient of a cylindrical wire as

$$30 + \frac{9}{\sqrt{a}}\text{ kV./cm.} \qquad . \quad . \quad . \quad (2\text{–}30)$$

for air at atmospheric pressure.

Continuing the consideration of the wire and cylinder type of gap, the stress at the surface of the wire is

$$X_s = V_p/a \log_e (A/a)$$

where V_p is the potential difference between wire and cylinder and A is the cylinder radius. If local breakdown occurs it may be assumed that the wire becomes covered with a conducting layer of disrupted gas, the effective radius increasing from a to $a + da$ where da is the thickness of the disrupted layer. Hence

$$\frac{dX_s}{da} = \frac{V_p}{(a \log_e A/a)^2}[1 - \log_e (A/a)] \qquad . \quad . \quad (2\text{–}31)$$

From this it will be seen that, if $A/a > \varepsilon$, (2–31) is negative and so the first effect of an applied voltage is to reduce the stress at the wire. If $A/a < \varepsilon$, (2–31) is positive and the stress is increased. In the first instance, corona occurs, and in the second a complete breakdown. Of course, even in the first instance, if V_p is increased sufficiently the effective value of a will ultimately become such that $A/a < \varepsilon$.

POINT AND PLANE

If the end of a fine wire be regarded as a sphere whose radius is equal to that of the wire, the breakdown gradient may be found in a similar manner to that for a wire and cylinder. The field strength at any radius r is

$$X = X_s \frac{a^2}{r^2}$$

and the potential corresponding to X

$$V = \int_a^r X dr = X_s a^2 \int_a^r \frac{dr}{r^2} = X_s a^2 \left(\frac{1}{a} - \frac{1}{r}\right) \qquad . \quad (2\text{–}32)$$

Plotting (2–32) under the breakdown curve for normal fields will again enable X_s to be found.

The foregoing treatment can, of course, be applied to only relatively simple discharge systems on account of the difficulty of finding simple expressions for the electrostatic field. In the case of two parallel wires widely separated, the field at each is practically radial and hence the field conditions at the wires are similar to those of the wire and cylinder. Hence it is to be expected that the breakdown surface gradient will be given by an expression similar to (2–30) and this actually is so, Peek giving

$$X_s = 30\left(1 + \frac{0{\cdot}301}{\sqrt{a}}\right) \text{kV/cm.} \qquad . \quad . \quad (2\text{–}33)$$

EFFECT OF GAS DENSITY

The effect of varying the gas density and pressure with non-uniform fields is similar to that for the uniform case following the principle of similarity. Thus (2–33) may be written

$$X_s = 30p\left(1 + \frac{0{\cdot}301}{\sqrt{ap}}\right) \text{kV/cm.}$$

while Whitehead gives the following formulae

Wire and cylinder $31p\left(1 + \frac{0{\cdot}308}{\sqrt{ap}}\right)$ kV/cm.

Spheres $27{\cdot}2p\left(1 + \frac{0{\cdot}54}{\sqrt{ap}}\right)$ kV/cm.

where p is in atmospheres.

The Spark Discharge

When breakdown occurs in a gas, the pressure of which is of the order of an atmosphere, the discharge initially takes the well-known form of a spark. Whether other states follow this depends on the nature of the circuit to which the electrodes are connected. However, although the duration of a spark is of the order of micro-seconds, it is very probable that even in this brief time the current may reach "arc" values as, indeed, is indicated by (2–16). The conditions of breakdown are given by this equation, and hence also of sparking. (2–16), however, indicates that having produced sparking conditions at a discharge system a spark immediately follows. In practice this is not so, for it is common knowledge that a higher voltage is needed to cause the first spark than subsequent sparks, provided the latter follow the first with sufficient rapidity. Also it is well known that the breakdown potential of a spark-gap is higher if the voltage is rapidly applied than if slowly.

The foregoing statements indicate that a time-lag exists between the application of an impulse voltage and the subsequent breakdown of a discharge system. This time-lag consists of two separate components: (1) a statistical time-lag necessitated by the time required for casual electrons to appear in the gap to initiate the discharge; (2) a formative time-lag necessitated by the time for the development of ionization and the progress of the discharge across the gap. The ratio of the breakdown figure when the voltage is slowly applied to that when the application is rapid is termed the impulse ratio of a spark-gap. Fig. 2–10 shows the manner in which the breakdown voltage increases with decreasing time of voltage application for points in air at atmospheric pressure.

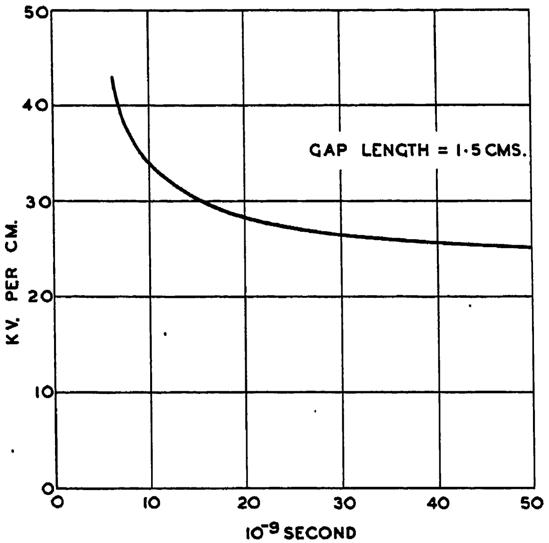


Fig. 2–10

Formative time-lag is definitely in evidence whenever corona precedes breakdown. In these cases the average breakdown gradient is less than that for a uniform field with the same electrode spacing. As has been previously shown, where corona occurs we have a virtual increase in the dimensions of the electrode, with the result that the space-charges increase the gradient in the gap up to the breakdown point. Referring to Figs. 2–11 and 2–12,* it may be concluded that where the average breakdown gradient is less than

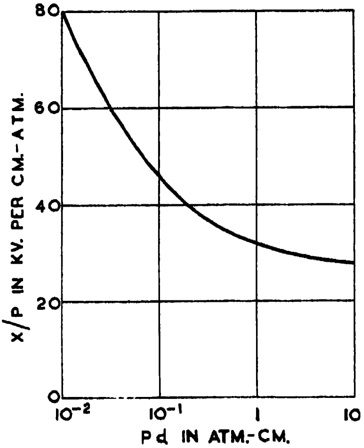


Fig. 2–11

---

*Fig. 2–11 refers to uniform fields and Fig. 2–12 to points of 0·4 cm. separation.

that for uniform fields, the field has been distorted by space charges.

In order to eliminate the statistical time-lag and study the period of the formative time-lag, initiating electrons may be created in the field of the spark-gap. This is usually effected by means of radioactive substances or else by ultra-violet radiation. Ultra-violet radiation may be produced by several methods, a common method being the employment of a mercury-vapour lamp. If a gap is irradiated by means of the latter, the majority of the photo-electrons are produced from the cathode metal rather than the gas. This may be proved by altering the angle of the incident radiation. The photo-electric current at the cathode is of the order of 10^{-12} amp. cm.^{2}, this corresponding to

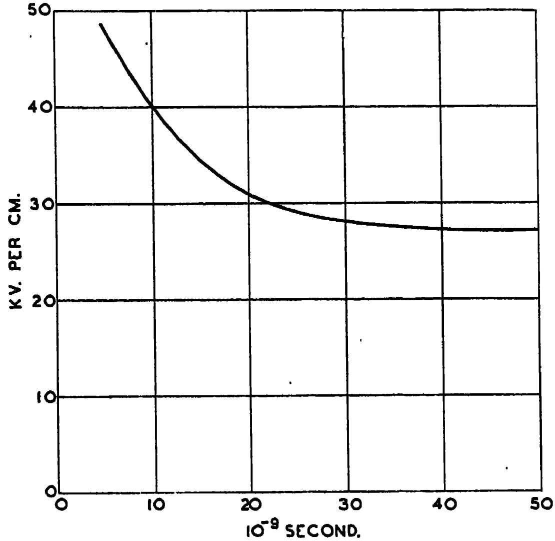


FIG. 2-12

$$\frac{10^{-12}}{1 \cdot 59 \times 10^{-19}} = \text{electrons per cm.}^2 \text{ per sec.}$$

or 6 electrons per cm.^{2} per microsec. approx.

Hence, in this case, statistical time-lag would probably not occur, providing the static breakdown voltage were not applied for a time shorter than 1/6 microsec. and the cathode area were not less than

1 cm.[2] The importance of the cathode area in the reduction of statistical time-lag is at once evident, for it is clear that the greater this area the greater will be the number of photo-electric initiating electrons available. This explains why gaps with a large area (whether the initiating electrons are produced by ultra-violet radiation or some other means) such as spheres, wire and cylinder, etc., have low impulse ratios compared with gaps formed by points. We see, then, that the criterion for statistical time-lag is the rate of application of voltage, the number of initiating electrons produced per square centimetre per second, and the gap or cathode area depending on whether the electrons are produced in the former or at the latter. Where the electrons are formed in the gap it is, of course, the volume of this which is of importance rather than the area.

A method of eliminating statistical time-lag is by means of the ultra-violet radiation of a neighbouring spark or corona discharge. With regard to the spark, experiments indicate that cathode photo-electric currents of the order of 10^{-6} amp. per cm.^{2} may be produced by this means, i.e. 10^6 times as great as where the cathode is irradiated by means of a mercury-vapour lamp. With regard to the corona discharge, a well-known application of this to reduce impulse ratio is found in the third point spark-gap,* extensively used for magneto and ignition-coil testing. In this gap, which may consist of points or spheres, a third pointed electrode is situated close to but not actually touching one of the main electrodes. The third point is either connected to the remote electrode by means of a high resistance of the order of 40 megohms, or left isolated. The corona discharge which occurs at the point thus furnishes a photo-electric supply of electrons to the main gap.

Summarizing, we may say that the time-order of statistical time-lag is that of the inverse rate of formation of photo-electrons at the cathode by the source of radiation.

FORMATIVE TIME-LAG

We have already seen that providing an adequate photo-electric current is produced, the statistical time-lag may be eliminated. When this is done, experiments show that the formative time-lag is of the same order as that of the time taken by the electrons to cross the gap. Referring to Figs. 2–13 and 2–14, which are reproductions of oscillograms of spark discharges, it will be seen that after the static breakdown voltage is reached breakdown occurs within about 2 × 10^{-7} sec. By applying overvoltages to the gap, this time may be still further reduced. A further method of studying formative time-lag consists of photographing the luminosity of a spark

---

* *Aircraft Electrical Engineering*, F. G. Spreadbury (Pitman).

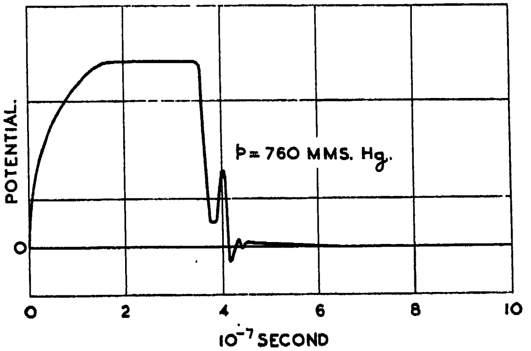


FIG. 2–13

as it develops. Typical photographs show that a luminous streamer commences at the cathode and ultimately, after a period of the order of 10^{-8} sec., terminates at the anode. An important discovery

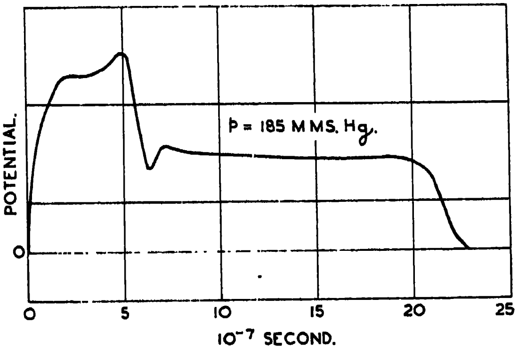


FIG. 2–14

of the photography is that for high values of pd a second streamer tends to occur, originating in the gap or close to the anode, eventually joining up with the streamer from the cathode.

LOWERING OF SPARK POTENTIAL BY RADIATION

An effect of irradiation of a spark-gap is a lowering of the breakdown potential. This effect becomes in evidence when the photoelectric current at the cathode exceeds about 10^{-11} amp. per cm.^{2}

Hence it is particularly noticeable when the gap is irradiated by a neighbouring spark. As previously stated, under these circumstances i_0 may be as large as 10^{-6} amp. per cm.^{2} and for such currents a large positive space-charge may build up in the gap due to the relative mobilities of the positive ions and electrons. The resulting distortion of the field may then be such that the value of α (the first Townsend coefficient) is higher than for the externally applied field alone. Hence, in these circumstances, the gap will break down for a lower value of X than when i_0 is negligible.

The Streamer Theory of the Spark Discharge

The previously given theory of breakdown and the spark discharge is largely due to Townsend and has been generally accepted for the past forty years or so. During the past few years, however, certain observations have indicated that the Townsend theory is probably true for spark-gaps only where pd does not exceed about 200, p being expressed in millimetres Hg, and d in centimetres. According to the Townsend theory, the criterion of breakdown is $\gamma(e^{\alpha d} - 1) \geq 1$, or, if $e^{\alpha d}$ is large compared with unity, $\gamma e^{\alpha d} \geq 1$. Now, $e^{\alpha d}$ is the number of positive ions generated by each electron in its passage from cathode to anode, and γ is the probability of a positive ion subsequently ejecting a secondary electron from the cathode. It is evident that if breakdown is to occur, the formative time-lag must be of the time-order of the positive ions to cross the gap from anode to cathode. If we take the mobility of an ion to be that determined on page 61, then, assuming a breakdown potential of 30,000 volts per cm. and a 1 cm. gap, the formative time-lag is of the order of 10^{-5} sec. In dealing with the glow discharge, it will be shown that the velocity of electrons to that of ions is at least 100 : 1; hence it follows the time-order of electrons in crossing a gap as above must be 10^{-7} sec. For gaps longer than 1 cm. the formative time-lag must, of course, be greater than 10^{-5} sec.

Prior to 1925, little information was available on spark time-lags as facilities did not exist for their measurement. Ultimate measurements have revealed, however, that for sparks occurring at atmospheric pressure the formative time-lag is of the order of 10^{-7} sec. rather than 10^{-5} sec. Nevertheless, measurements with sparks occurring at values of pd up to about 200 give time-lags of 10^{-5} sec. and upwards, thus confirming the Townsend theory in this region. For sparks occurring at $pd > 200$ it is evident that the formative time is such as to preclude the movement of positive ions in the gap. Thus, in these cases, the spark must develop through electron movement only.

A further difficulty with the Townsend theory for gaps having $pd > 200$ is that the sparking potential is largely independent of cathode material. As shown on page 66, the breakdown potential should be dependent on γ and this quantity dependent on the cathode material. (See page 626.) Other difficulties are as follows—

Gaseous discharges exist where no cathode phenomena occur. Such are positive-point corona and the lightning discharge.

Cloud-track photographs of the spark discharge and photographs of its luminosity show that in its initial phases the spark possesses streamers which originate at the anode or, in the case of overvolted gaps, at the anode and in the gap itself. Such streamers only commence to exist for $pd > 200$.

The foregoing difficulties in the complete applicability of the Townsend theory clearly indicate the necessity of some other mechanism to explain the spark discharge for high pressures. The mechanism must involve the following—

(1) Electron movements only must occur, the positive ions remaining relatively stationary.

(2) The process must be independent of the cathode and depend on processes within the gas.

(3) Ionization must proceed from the anode and/or gap region towards the cathode.

(4) Due to the filamentary character of a spark, it must be initiated by a single electron along a narrow path.

THE STREAMER THEORY

A theory covering the foregoing necessities of the spark discharge has been developed, principally by Raether, Loeb, and Meek, the quantitative aspect being due to the last-named. Considering a parallel-plane gap 1 cm. in length, let it be assumed that the cathode is under the influence of ultra-violet radiation to the extent that 1 electron per cm.^{2} per microsec. is ejected from its surface. Further, it will be assumed that the potential difference across the gap is 31,600 volts and the gas air at normal pressure. Hence $X/p = 41{\cdot}6$ volts/cm. per mm. Hg.

Considering an electron leaving the cathode, it rapidly acquires a drift velocity of the order of 10^7 cm. per sec., this velocity actually having been measured by Raether. During its passage across the gap the electron creates further electrons at the rate of α per cm., so that in a distance x from the cathode the total number of electrons created is $\varepsilon^{\alpha x}$. This rapidly increasing quantity is appropriately termed an *electron avalanche*. Behind the avalanche are $\varepsilon^{\alpha x}$ positive ions which virtually remain where formed during the time of 10^{-7} sec. taken by the electrons to cross the gap. During the advance of the avalanche its head is spreading laterally, due to the random diffusion of the electrons. According to Raether, the average radial distance of diffusion is given by $r = \sqrt{2Dt}$, where $t = x/v$, v being the electron velocity and D a coefficient of diffusion. Up to this state of affairs we have the Townsend mechanism represented by (2–3), for the positive ions are in no position to effect electron multiplication. Hence an avalanche which has crossed the gap does not constitute a breakdown.

The positive ion density will, of course, be a maximum at the anode. The ions will form a cone, the base of which will be situated at the anode and the apex approximately at the cathode. Thus, a considerable positive space-charge will exist near the anode, giving rise to a field of some magnitude. In order to calculate the strength of this field it is assumed that the ions are contained within a spherical volume at the head of the avalanche instead of the actual conical volume. The radius of the sphere is taken to be $r = \sqrt{2Dt}$ and the field due to the space-charge is then given by

$$X_s = 4\pi ne/4\pi r^2 = ne/r^2$$

where n is the number of ions in the sphere. If N is the ion density, then $n = 4\pi r^3 N/3$ and $X_s = 4\pi rNe/3$. In a distance dx at the end of a path x the number of ions formed is $\alpha\varepsilon^{\alpha x}dx$ and

$$N = \frac{\alpha\varepsilon^{\alpha x}dx}{\pi r^2 dx} = \frac{\alpha\varepsilon^{\alpha x}}{\pi r^2} \quad . \quad . \quad . \quad (2\text{–}34)$$

Hence

$$X_s = \frac{4}{3} e\alpha\varepsilon^{\alpha x}/r$$

or

$$X_s = \frac{4e\alpha\varepsilon^{\alpha x}}{3\sqrt{2Dt}}$$

For the case under consideration, $\alpha = 17$ and r, as observed by Raether, is 0·013 cm. This gives X_s as 6000 volts per cm. and X_s/X as 0·20. Thus the field due to the space-charge is 20 per cent of that near the cathode.

In addition to ionizing the gas molecules, the electrons also produce a large number of excited atoms and molecules. These excited atoms and molecules produce ultra-violet radiation which liberates from the gas and cathode photo-electrons, which in turn create further ionization. The photo-electrons will, of course, create further avalanches, those in the gas being relatively short while those created at or near the cathode will be long and similar to the original avalanche. Now those photo-electrons created near the space-charge channel of positive ions, particularly those near the anode, are in an enhanced field, which exerts a directive action, drawing them within the field. The electrons resulting from ionization due to photo-electron avalanches in the combined fields X_s and X, pass into the positive space-charge, thus rendering it a conducting plasma, which commences at the anode. The positive ions left behind by each photo-electron avalanche have the effect of extending the positive space-charge towards the cathode, because the applied field tends to distribute itself between the cathode and the positive space-charge rather than between cathode and anode. Thus the positive space-charge develops towards the cathode from the anode as a self-propagating positive space-charge streamer.

As previously stated, streamers as just described have been observed and photographed, this tending to confirm the foregoing theory. The velocity of propagation of a streamer is extremely rapid, as it depends on photo-ionization, photon propagation at the velocity of light, and short-distance motion of electrons in high fields near the space-charge. In one case observed by Raether the velocity of propagation was of the order of 10^8 cm. per sec. As the streamer approaches the cathode, the potential of the gap is mainly concentrated between the cathode and the edge of the streamer due to the relatively high conductivity of the plasma. Thus the value of α outside the plasma will be greatly increased, leading to intense ionization and the possibility of high secondary emission from the cathode. When the streamer reaches the cathode, the gap is bridged by a conducting filament and hence the gap voltage collapses to a low figure. Should a cathode spot form then, unless the current is limited by an external resistance, an arc will develop.

NECESSARY CONDITIONS FOR STREAMER FORMATION

In order that a streamer may form and lead to gap breakdown, it is necessary that sufficient excited gas molecules are formed in the space-charge channel to produce an adequate supply of photo-electrons. This means that a certain space-charge density is necessary if streamer formation is to occur. Again, a certain space-charge density is essential to ensure that sufficient photo-electrons are drawn into the space-charge to form further avalanches to advance the streamer tip. We see, therefore, that the criterion for gap breakdown is an adequate space-charge density to ensure, firstly, the necessary density of excited molecules for photo-ionization, and, secondly, to create the necessary field to draw photo-electrons into the space-charge channel. If the space-charge channel is broad, then streamer formation is unlikely, as the necessary space-charge density may not be obtained. Thus, streamer formation is unlikely to occur as low gas densities.

We have already seen that the relation between X_s and N may be written

$$X_s = 4\pi r e N/3$$

and

$$N = \frac{\alpha\varepsilon^{\alpha x}}{\pi r^2}$$

As N and X_s are proportional to each other the criterion for streamer formation may be expressed either in terms of X_s or N. The criterion given by Meek for streamer formation is that $X_s = X$, i.e. the field due to the space-charge must equal the external impressed field. From this Meek finds the condition for breakdown to be

$$X_s \fallingdotseq 5{\cdot}27 \times 10^{-7} \frac{\alpha\varepsilon^{\alpha x}}{(x/p)^{\frac{1}{2}}} \text{ volts/cm.}$$

For further details of the streamer theory of the spark discharge the reader is referred to Loeb and Meek's book* on this subject.

The Glow Discharge

The conditions immediately succeeding breakdown of a gas largely depend on the circuit of which the electrode system forms a part and the gas pressure surrounding the electrodes. In general, we may distinguish four states of gaseous conduction: (1) the pre-breakdown state with its relatively small current; (2) the disruptive discharge (spark); (3) the glow discharge, and (4) the arc discharge. Where the power is strictly limited, such as, for example, with an induction coil or magneto discharge, it is extremely probable that all four states occur, following each other in rapid succession in the above order. If following state (2) the circuit power is adequate and the electrode current density is maintained below a certain value, state (3), the glow discharge, occurs. A most conspicuous feature of this discharge is its luminosity, which, it may be mentioned, forms the basis of what are known as cold cathode luminous discharge tubes (neon signs, etc.).

* *The Mechanism of the Electric Spark* (Oxford University Press).

After a glow discharge is initiated it is found that the voltage necessary for its maintenance is invariably less than that necessary for its initiation, the ratio of these voltages depending on the gas pressure and the electrode spacing. Now there is no reason for supposing that the energy with which the positive ions strike the cathode is less after the discharge has commenced than at its initiation. Actually this energy should be the same for the continued liberation of electrons to maintain the discharge. Thus, as the average voltage gradient across the gas is less after conduction than before, its distribution must be different. The change of distribution is due to the presence of what are known as space-charges within the gas. Due to the higher mobility and longer mean free path of the electrons as compared with those of the positive ions, the density of the latter is far greater than that of the electrons. Thus, there is in general a net positive charge in the conducting space, the field distribution no longer being solely determined by the surface charges on the electrodes but by the space-charges in the gas as well.

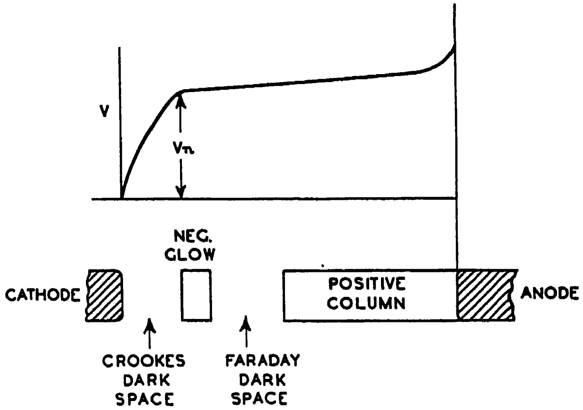


Fig 2–15

The glow discharge is so termed because of its characteristic appearance. Although it may be formed between short gaps (a millimetre or so) at atmospheric pressure,* for demonstration purposes more satisfactory results are obtained between wide gaps at pressures considerably below that of the atmosphere. Under the latter circumstances it is readily observed that the glow is divided by a number of dark spaces. Furthermore, the potential distribution along the discharge is related to these spaces. This relationship is shown by Fig. 2–15, from which the outstanding importance of the Crookes dark space is evident. It is clear that a large proportion of the applied voltage is absorbed by this space, the amount being termed the cathode fall. It may be defined as the potential difference between the cathode and the point of minimum field strength, i.e. least potential gradient. This potential difference is independent of the current strength and gas pressure and depends only on the nature of the gas and of the cathode metal. In general it is approximately equal to the minimum sparking potential of the gas and thus corresponds to the minimum potential at which the discharge can be maintained.

**Aircraft Electrical Engineering*, p. 15.

THE POTENTIAL CHARACTERISTIC

An explanation of the potential distribution along a glow discharge may be found in the following manner. Due to the relationship

$$eV = eXl = \tfrac{1}{2}mv^2$$

an electron acquires a velocity $\sqrt{1840A}$ times as great as a positive ion of atomic weight A in a free fall through the same potential difference $V = Xl$. Due to this relatively high velocity, the ions, relatively to the electrons, may be considered at rest. Hence, referring to (1–42) and (1–45), $\overline{C_r} = \bar{C}$ and

$$l_e = \frac{4}{\pi(\sigma_e + \sigma)^2 n} \qquad . \quad . \quad . \quad (2\text{–}35)$$

where σ_e is the diameter of the electron and l_e is the electronic mean free path. The velocity of the ions is of the order of magnitude of the gas-kinetic velocity so that

$$l_i = \frac{1}{\sqrt{2}\pi\sigma^2 n} \qquad . \quad . \quad . \quad . \quad (2\text{–}36)$$

where l_i is the ionic mean free path. As σ_e is negligible compared with σ we have

$$\frac{l_e}{l_i} = 4\sqrt{2}$$

and in a given field the velocity of the electrons is to that of the ions as

$$\sqrt{1840A \,.\, 4\sqrt{2}} : 1 \text{ or as } 102\sqrt{A} : 1 \quad . \quad . \quad (2\text{–}37)$$

Now the positive column or plasma is characterized by linear potential rise, which means that here $d^2V/dx^2 = 0$, i.e. this region is free from space-charge. Thus, within the plasma

$$n_+ = n_-$$

also

$$i = i_+ + i_- = en_+v_+ + en_-v_-$$

so that

$$\frac{i_-}{i_+} = \frac{v_-}{v_+} = 102\sqrt{A} : 1$$

Hence in regions of no space-charge, in particular the positive column, the current carried by electrons is at least one hundred times that carried by positive ions.

At the cathode and anode the current is mainly carried by positive ions and electrons respectively, i.e.

$$i = en_+v_+ = en_-v_-$$

If the field is the same at both electrodes, then the ratio of positive to negative space-charge before cathode and anode is

$$\frac{en_+}{en_-} = \frac{v_-}{v_+} = 102\sqrt{A} : 1$$

Now

$$\begin{aligned} \frac{d^2V}{dx^2} &= -4\pi\rho \\ &= -102\sqrt{A} \,.\, 4\pi en_- \quad \text{in cathode region} \\ &= 4\pi en_- \quad \text{in anode region} \end{aligned}$$

Thus, we see that the potential characteristic must turn sharply downward at the cathode and slightly upward at the anode.

THE NEGATIVE GLOW

If the gas-filling of a glow-discharge tube is air, the negative glow is blue and the positive column pink. Visual inspection of the cathode shows that while the cathode fall is independent of current ("normal" cathode fall) the cross-section of the negative glow is less than that of the cathode and is proportional to the current. From this it may be concluded that the normal cathode fall, V_n, corresponds to a region of constant current density, J_n. Accompanying these conditions is a normal thickness of Crookes dark space d_n. For a given gas at density p, pd_n is constant. Actually this is because V_n is constant, for according to Paschen's Law, V_n

is a function of pd_n. Hence while V_n remains constant so also must pd_n and $d_n \propto 1/p$. It is also found that

$$\frac{J_n}{p^2} = \text{constant} \quad . \quad . \quad . \quad (2\text{–}38)$$

and from this it is evident that for a given current the cross-section of the negative glow varies inversely as p^2.

When the current rises above the value at which the cathode is covered by the negative glow, the cathode fall no longer remains constant but commences to rise. Also d_n becomes reduced, the various quantities being related by

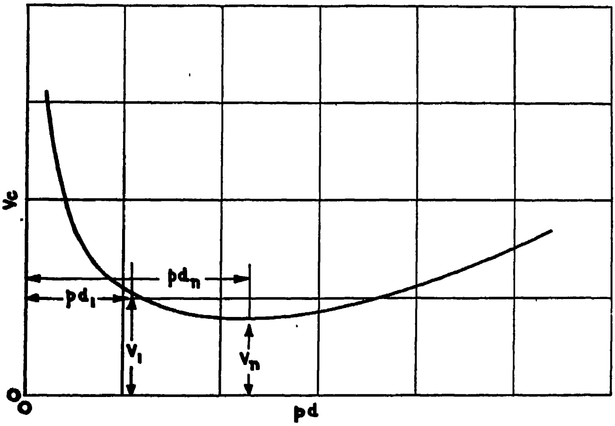


FIG. 2–16

$$\frac{Jpd^3}{V^2} = \text{constant} \quad . \quad . \quad . \quad (2\text{–}39)$$

The decrease in d as V_c increases above V_n is explainable if the positive column and anode drops are ignored and the cathode fall is assumed to be equal to the striking potential of the tube. In this case the cathode fall must be on the breakdown curve, normally at the point of minimum striking potential, as shown by Fig. 2–16. If, now, V_c increases while p remains constant, we see from Fig. 2–16 that pd must decrease, i.e. d must fall to enable V_c to move up the left-hand branch of the curve.

It is possible to deduce (2–38) and (2–39) in the following manner. From page 87

$$J = en_+v_+ = \rho_+v_+$$

and $$\rho = J/v_+ \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}40)$$

Also $$v_+ = aX/p \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}41)$$

and from Poisson's Law

$$\frac{d^2V}{dx^2} = \frac{dX}{dx} = 4\pi\rho \quad . \quad . \quad . \quad . \quad (2\text{–}42)$$

From (2–40), (2–41), and (2–42)

$$\frac{dX}{dx} = A\,\frac{Jp}{X}, \text{ where } A = \frac{4\pi}{a}$$

Integrating

$$X^2 = 2AJpx$$

$$X = \sqrt{2AJp}\,x^{\frac{1}{2}}$$

Now

$$V = \int_0^d X dx = \sqrt{2AJp}\int_0^d x^{\frac{1}{2}}\,dx$$

$$= \tfrac{2}{3}\sqrt{2AJp}\,d^{\frac{3}{2}}$$

and

$$V^2 = \tfrac{8}{9}\,AJpd^3 \quad . \quad . \quad . \quad . \quad . \quad (2\text{–}43)$$

which leads to (2–39).

Multiplying (2–43) by p^2 and re-arranging, we have

$$\frac{V^2}{p^3d^3} = \frac{8A}{9}\cdot\frac{J}{p^2} \quad . \quad . \quad . \quad (2\text{–}44)$$

and while V and d are, respectively, equal to V_n and d_n, pd is constant and (2–44) is equivalent to (2–38).

THE CROOKES AND FARADAY DARK SPACES

The electrons emitted from the cathode surface are few in number and possess low initial velocities. Due to the high field strength in the vicinity of the cathode, the electrons rapidly gain velocity and energy, but because of this their excitation and ionization probability is small. Consequently, little light is emitted in the region of the cathode fall, this giving rise to the Crookes dark space.

At the edge of the cathode fall space the net space-charge density falls to zero or may even become negative. This reduces the velocity of the electrons whose excitation and ionization probabilities, in consequence, rise, causing the negative glow. In the latter the velocity becomes so low that the excitation and ionization probabilities fall, giving rise to the Faraday dark space. In this space few positive ions exist, the space charge being negative. This accelerates the electrons, which ultimately produce further excitation and ionization, the latter marking the commencement of the plasma or positive column.

PLASMA CHARACTERISTICS

The plasma, as we have already seen, is a region of weak fields where the ionic and electronic concentrations are equal, giving rise to a zero net space-charge. Due to the uniform field, ionization occurs equally along the positive column, resulting in its uniformly luminous appearance. As there is no space-charge, the current in the plasma is not space-charge limited, with the result that large currents may exist for small potential differences. In addition to the electrons which drift towards the anode (thus constituting the conduction current), random ion and electron currents also exist within the plasma. Radial ion and electron currents also occur, these resulting in a loss of ions and electrons from the plasma to the walls of the discharge tube. This loss is termed de-ionization and, of course, necessitates ionization to occur constantly within the plasma to repair the loss. Because of the higher mobility of electrons, they diffuse to the walls of the tube more rapidly than the ions, with the result that the walls are generally a few volts negative with respect to the plasma. The potential variation at the walls may be regarded as a boundary between the latter and the plasma, and is termed a sheath. In general, sheaths always exist between the electrodes (as well as the walls) of gaseous discharge tubes and the plasma, and the cathode dark space in a glow discharge may be regarded as a sheath.

THE ELECTRODES

It is evident from Fig. 2–15 that the greatest potential falls occur in the immediate vicinity of the electrodes. Hence the rate of energy dissipation and temperature rise are a maximum at these situations. The temperature is, of course, greatest at the cathode, and it is clear that electrode design calls for careful consideration if undue temperatures are to be avoided. A further detrimental cathode effect is known as sputtering, and consists of progressive disintegration of the electrode due to positive ion bombardment. The electrode particles are deposited on the walls of the tube, thus tending to absorb the luminosity of the tube. In addition, the particles take up some of the gas during deposition, with the result that a process known as clean-up occurs, with a lowering of tube gas pressure and a consequent modification of tube performance. It is evident that electrodes and gas fillings are desirable, which give a low cathode fall, for this leads to low temperatures and long electrode life. Table 19–2 gives some typical figures for various gases and electrode materials. Reference will again be made to this subject when dealing with luminous discharge tubes.

The Arc Discharge

If the current in a glow discharge is increased beyond the region of normal current density, an abnormal cathode fall and constricted dark space result. If the process is continued, a sudden transition ultimately occurs in which the discharge voltage falls to a low value, accompanied by an increase in current with the concentration of the latter on to a relatively small area of the cathode. When the foregoing phenomena occur, the current is almost entirely controlled by the circuit to which the discharge electrodes are connected. The latter form of discharge was termed an arc by Sir Humphry Davy, on account of its tendency (due to convection currents) to form the geometrical figure of that name when formed between horizontal electrodes.

Although the magnitudes of the currents normally associated with glow and arc discharge differ widely, both forms have much in common, namely a cathode and anode fall and plasma. The principal difference is the mechanism by which electrons are released from the cathode. With the arc they are released in such copious quantities that they largely neutralize the positive space-charge associated with the glow, thus giving rise to a low cathode fall, of the order of 10 volts, which accounts for the relatively low voltage usually accompanying an arc discharge.

Due to lack of sufficient evidence, complete agreement does not exist regarding the method by which electrons are released from the cathode in an arc discharge. The most obvious explanation is that they are released thermionically. As the current is increased in a glow discharge the rate of energy dissipation at the cathode continually increases, particularly under abnormal conditions. Hence the cathode temperature may increase to a value at which thermionic emission occurs, thus releasing electrons at such a rate as to neutralize the positive space-charge. Objections to the foregoing explanation are as follows: An arc may be struck and maintained in constant motion over the electrodes by means of a radial magnetic field. By this process the cathode is maintained at a temperature below which thermionic emission does not occur. Slepian estimated that the transition from glow to arc occurred for a current density of 5 amps. per sq. cm. He calculated that the thermal conductivity of a copper cathode is such that it cannot with this current density attain a sufficient temperature for emission to occur. Furthermore, he states that, in any case, the necessary temperature would take several seconds to attain, whereas an arc may be struck in a period as short as 10^{-5} sec.

In the case of a mercury arc it has been considered that the cathode fall energy may be retained in a layer of mercury only 10^{-6} cm. thick. Should this be the case, then a temperature of 3000° C. might be attained within a few seconds.

In view of the somewhat unsatisfactory state of the thermionic theory of the arc discharge it was suggested by Langmuir that the electrons might be emitted from the cathode by field or auto-electronic emission. This, of course, would not necessitate a hot cathode. Again, however, it has been objected that the current densities necessary to create the requisite field strength are far in excess of those actually needed to produce an arc.

The Mercury Arc

Two forms of mercury arc exist: that occurring in mercury vapour only and that where the vapour is in the presence of its liquid. An arc occurring in the latter has a mercury pool for cathode from which the electron emission is derived. The exact method by which the electrons are emitted from the pool is as obscure as with the arcs previously described. The electrons are derived from what is known as the cathode spot, i.e. a small highly-luminous rapidly-moving spot on the surface of the mercury pool. From the spot proceeds a blast of mercury vapour having a velocity of the order of 10^5 cm. per sec. As this blast deflects the positive ions proceeding to the cathode, the ions are ever seeking fresh points of contact, causing the spot to wander over the surface of the mercury pool in a highly erratic manner. The cathode spot current density is estimated at 4000 amp. per cm.^{2} and its temperature as 2400° K.

Unlike the glow, which has a constant or slightly rising voltage-current characteristic, the arc has a falling characteristic, i.e. dV/dI is negative. For currents in excess of one ampere the voltage/current relation may be expressed by

$$V = A + BI^{-n} \qquad . \quad . \quad . \quad (2\text{–}45)$$

where A and B are functions of the arc length. According to Nottingham the value of n is proportional to the boiling temperature of the anode. Where the ambient gas is air, n corresponds to the boiling temperature of the oxide of the anode metal. Table 2–2 gives values of n for a number of metals when employed as anodes. It is evident that for sufficiently large currents the second term

TABLE 2–2

Anode Material	Exponent n
Tungsten	1·38
Platinum	1·15
Carbon	1·00
Lead	0·48
Copper	0·67
Aluminium	0·65
Nickel	0·64
Silver	0·625
Antimony	0·46
Bismuth	0·445
Zinc (oxide)	0·57
Zinc	0·345

of (2–45) may be negligible compared with A, in which case the arc voltage is approximately constant and independent of current strength. Table 2–3 gives some values of A and B for various arc lengths.

TABLE 2–3

	Cu in Air		W in Air		Fe in Air	
Length mm.	A	B	A	B	A	B
1	—	—	—	—	19·1	27·7
2	—	—	40·1	16·0	21·5	35·6
3	29·0	39·2	43·0	20·3	—	—
4	32·0	43·9	45·3	25·8	24·0	50·2
5	34·0	49·0	47·4	31·0	—	—
6	36·5	52·2	49·3	37·0	25·1	60·8
7	37·5	57·1	51·1	41·8	—	—
8	38·5	62·9	53·0	46·8	26·6	72·5
9	40·0	66·5	54·7	52·0	—	—
10	41·0	70·5	56·0	58·0	—	—
25	—	—	—	—	50·0	117·0
50	—	—	—	—	70·0	190·0

Probes and Probe Measurements

The distribution of potential in the various parts of the glow and arc discharges have so far been discussed without reference to any method of ascertaining this distribution. We shall now consider the usual method of obtaining such information. If an insulated conductor (termed a probe) is inserted in a discharge path, it will acquire some potential relative to that part of the discharge in which it is situated. While the conductor is insulated it follows that the net current to it is zero. Assuming it is in the plasma, where the concentration of ions and electrons is equal, the number of electrons arriving at the probe per second will initially be greater than the number of positive ions because of the greater mobility of the former. Hence in order that the net current to the probe may be zero, the probe takes up a negative potential relatively to the plasma so that it tends to repel electrons and attract positive ions in order that the number of each reaching the probe in a given time shall be the same. This condition, of course, corresponds to zero current and the probe is said to "float" and possess a "floating" potential.

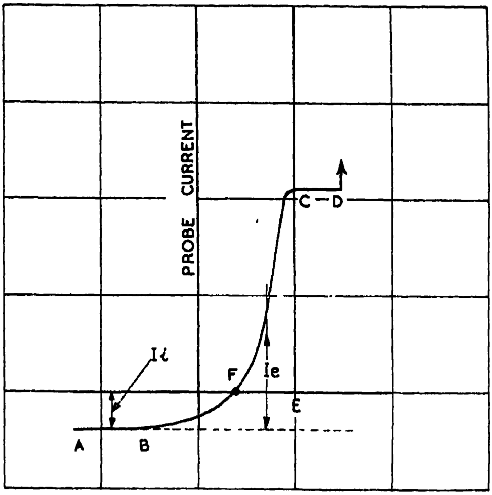


FIG. 2–17

If now, instead of allowing the probe to float, a potential difference is maintained between it and either the cathode or anode, a probe current/voltage characteristic may be obtained. It is found that this takes the form shown by Fig. 2–17. While the probe is largely negative with respect to the plasma, the branch of the curve *AB* is obtained. Here the electrons are repelled and the positive ions attracted so that the region *AB* is that for which only a positive ion current exists. This current is mainly due to the random motion of diffusion of the ions, which hence arrive at the probe under their own natural velocities which, incidentally, follow a Maxwellian law of distribution. Because the current is due to random motion, *AB* is a region of saturation. Now, if we commence to make the potential of the probe less negative, electrons will pass to the probe and the difference between the electron and positive ion currents will be collected. Hence the true electron current may be deduced from the branch of the curve *BC* by adding the constant value of the positive ion current, I_i, to the ordinates of the branch *BC*. At a potential equal to the floating potential, equal numbers of electrons and ions will be collected with the result that the ammeter in the external probe circuit will indicate zero. This, of course, corresponds to the point *F*. As the probe potential is increased beyond *F*, electrons are collected at an increasing rate because the electron-retarding action of the probe potential with respect to the plasma is decreasing. When the probe is at the same potential as the plasma, the probe will neither repel nor attract electrons and ions, which will diffuse to the probe under the action of their random motion. If, now, the potential of the probe is made positive with respect to the plasma there will be no further increase in electron current, but a decrease in positive ion current due to repulsion of the ions. Hence the only further possible increase in probe current is an amount equal to I_i, after which a pure electron saturation current results for further increases in probe potential. Reference to Fig. 2–17 therefore shows that when the probe potential is the same as that of the plasma, a point of inflexion occurs in the current/voltage characteristic just preceding saturation conditions. This point gives the potential of the plasma.* If the probe voltage is made more and more positive the main current may transfer to it, and the probe then becomes an auxiliary anode as indicated by *D*.

PROBE MEASUREMENTS

We have already seen that by means of a probe, employed in the manner described above, the floating and plasma potentials may be readily measured. In addition to these, various other quantities may be measured. For example, the number of particles which, possessing a Maxwellian distribution of velocity, pass through each square centimetre of a plane per second, is, from (4–30),

$$n\sqrt{\frac{kT}{2\pi m}}$$

Thus the random electron or positive ion current density is given by

$$J_r = ne\sqrt{\frac{kT}{2\pi m}} \quad . \quad . \quad . \quad (2\text{–}46)$$

---

*The exact value of the plasma potential may be obtained from the ordinate $(I_i - I_e)$, where I_i and I_e are obtained from *AB* and *CD*.

Now, J_r is given by I/A, where I is the current determined from a curve such as that of Fig. 2–17, and A is the probe area. Hence we have

$$J_{ri} = I_i/A = n_i e\sqrt{\frac{kT}{2\pi m_i}}$$

$$J_{re} = I_e/A = n_e e\sqrt{\frac{kT}{2\pi m_e}}$$

$$n_i = \frac{I_i}{Ae}\sqrt{\frac{2\pi m_i}{kT}}$$

$$n_e = \frac{I_e}{Ae}\sqrt{\frac{2\pi m_e}{kT}}$$

where n_i and n_e are, respectively, the positive ion and electron concentrations within the plasma, m_i and m_e are the masses of an ion and electron, and I_i and I_e are the positive ion and electron currents determined from the two saturated portions of the curve of Fig. 2–17. Plasma ion concentrations found in practice vary from about 10^{10} to 10^{17} ions per cm.^{3}, depending on the gas pressure.

ELECTRON TEMPERATURE

On the assumption that the electron velocity distribution within the plasma is Maxwellian (and there is evidence that it is) the electron "temperature" may be found from probe measurements. This temperature is, of course, given by $\frac{1}{2}m_e C^2 = 3kT/2$. Now while the probe potential is negative with respect to the plasma, the only electrons penetrating the positive ion sheath around the probe are those whose velocity and kinetic energy are such that they can overcome potential drop across the sheath. According to (1–18) the number of electrons per cubic centimetre having velocities between u and $u + du$ is

$$\frac{n_e}{a\sqrt{\pi}}\varepsilon^{-\frac{u^2}{a^2}}du$$

and the number of such electrons passing through 1 cm.^{2} per sec. is

$$\frac{n_e}{a\sqrt{\pi}}\varepsilon^{-\frac{u^2}{a^2}}udu$$

The number of electrons capable of penetrating the sheath and reaching the probe are those the velocity of which is given by

$$\tfrac{1}{2}m_e u_s^2 = eE_s$$

where E_s is the sheath potential difference, i.e. the potential difference between the probe and the plasma. Hence the electron current to the probe is

$$\frac{n_e}{a\sqrt{\pi}}\int_{\sqrt{\frac{2eE_s}{m_e}}}^{\infty} \varepsilon^{-\frac{u^2}{a^2}}u\,du$$

$$= -\frac{an_e}{2\sqrt{\pi}}\left[\varepsilon^{-\frac{u^2}{a^2}}\right]_{\sqrt{\frac{2eE_s}{m}}}^{\infty}$$

$$= \frac{an_e}{2\sqrt{\pi}}\varepsilon^{-\frac{2eE_s}{m_e a^2}}$$

$$= n_e\sqrt{\frac{kT}{2\pi m_e}}\,\varepsilon^{-\frac{eE_s}{kT}}$$

or $$J_e = J_{re}\varepsilon^{-\frac{eE_s}{kT}}$$

where J_e is the electron current density to the probe. When $E_s = 0$, $J_e = J_{re}$, a result which, of course, has already been deduced above.

Taking the logarithms of both sides of the last expression there results

$$\log J_e = \log J_{re} - \frac{e}{kT}E_s$$

or $$\log I_e = \log AJ_{re} - \frac{e}{kT}E_s$$

which is the equation to a straight line having a slope equal to $-e/kT$. The electron current I_e can be deduced from Fig. 2–17, in the manner previously described, and hence T determined. The fact that a straight line is usually obtained in these experiments justifies the assumption of a Maxwellian distribution of electron velocities.

SHEATH THICKNESS

Corresponding to the branch AB of the curve of Fig. 2–17, a practically pure positive ion space-charge exists between the plasma and the probe. Through this space-charge pass current carriers of one sign only (positive ions), which are derived from a plane where the potential gradient is zero (the plasma face of the sheath), from which they start with negligible velocities. These positive ions pass freely to the probe because the sheath thickness is normally only a small fraction of the ionic mean free path. From these facts it is apparent that the conditions are similar to those existing in a diode carrying a space-charge limited electron current. Hence the space-charge equation (8–5) for parallel planes may be applied to the positive ion current density with appropriate modification, thus

$$J_i = \frac{2{\cdot}33 \times 10^{-6}}{\sqrt{1839M}} \frac{E_s^{3/2}}{s^2}$$

where s is the sheath thickness, M is the atomic weight of an ion, and J_i is the positive ion current density. But the latter quantity is independent of E_s and equal to J_{ri}. Hence

$$s = \frac{2{\cdot}33 \times 10^{-6}}{J_{ri}\sqrt{1839M}} E_s^{3/4}$$

which determines the sheath thickness. It will be noted that s increases with E_s and is zero when the probe is at the potential of the plasma. If the probe potential is raised above that of the plasma an electron sheath results. In this case J_{ri} must be replaced by J_{re} and $\sqrt{1839M}$ by unity.

In the foregoing theory of the probe it has been assumed that the area of the probe is constant. For large area probes this is justifiable, but not for those of small area. In the latter case the effect of the sheath is to increase the virtual area. Thus in the case of a fine wire, to the wire radius must be added the sheath thickness in determining the probe's effective area. Because of the increase in probe area, with an increase in E_s, I_i increases with E_s. Hence with a fine wire probe the branch AB of Fig. 2–17 will not be parallel to the axis, but will slope downwards.

REFERENCES

Conduction of Electricity through Gases, J. J. Thomson and G. P. Thomson (Macmillan).

The Mechanism of the Electric Spark, L. B. Loeb and J. M. Meek (Oxford).

"The Electric Spark in Air," J. M. Meek, *J.I.E.E.* Paper, 1942.

CHAPTER III

THE ELECTRON

Cathode Rays

REVERTING to the glow discharge, we shall now consider certain phenomena which occur if the gas pressure within the discharge tube is continually reduced. In previously discussing this discharge it was tacitly assumed that the gas pressure was such that the various phenomena, such as the dark spaces, glows, etc., were clearly visible. This pressure is of the order of 1 mm. of mercury. If now the pressure is further and continually reduced it is found that the positive column becomes smaller, receding towards the anode, while the Crookes dark space becomes enlarged. When the pressure is so reduced that the Crookes dark space reaches the walls of the tube, the glass fluoresces brilliantly, the colour being green for soda glass and blue for lead glass. Investigation showed that this fluorescence is due to something emitted from the cathode to which Goldstein gave the name of cathode rays. The rays were closely studied by Crookes and the following summarizes his results as well as those of later investigators.

1. The cathode rays leave the cathode normally to its surface, travelling in straight lines, and are negatively electrified. This latter property was verified by J. Perrin, who caused the rays to pass into a metal cup connected to an electroscope.
2. The rays behave as if possessed of inertia, being capable of deflecting obstacles and of turning a small mica mill-wheel placed in their path.
3. Substances struck by the rays are heated, and, should the rays be focused by a concave cathode, they are capable of raising small objects to incandescence.
4. They are deflected both by an electric and a magnetic field.
5. Their production and properties are entirely independent of the nature of the gas employed in the tube.
6. A substance struck by the rays emits radiation known as X-rays.

NATURE OF CATHODE RAYS

At one time considerable controversy existed concerning the nature of cathode rays. In this country Crookes advocated the view that the rays consisted of streams of electrified particles projected from the cathode, while the continental physicists supported Goldstein's theory that the rays were a form of wave motion. It may be noted that (1) and (2) are suggestive of Crookes view. The controversy was finally resolved by Sir J. J. Thomson, whose remarkable researches proved the rays to be electrically charged particles to which he gave the name of negative corpuscles. These corpuscles are now known under the name of electrons, the electron being one of the most important entities in the realm of physics.

Mass, Charge, and Velocity of the Electron

THOMSON'S EXPERIMENTS

The behaviour of an electron within a magnetic field will be first considered. Considering a small current element ids placed within a magnetic field of strength H, the force acting on this element is

$$Hids \quad . \quad . \quad . \quad . \quad (3\text{–}1)$$

But i may be written as dq/dt where q is the magnitude of the electric charge. Hence

$$Hids = H\frac{dq}{dt}ds = Hdq\frac{ds}{dt}$$

or, replacing dq by e and ds/dt by v, we have

$$Hids = Hev \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}2)$$

where e and v are, respectively, the charge and velocity of the electron. Now the force on the electron, given by (3–2), will act at right-angles both to its direction of motion and to the direction of the field. A body which experiences a force always perpendicular to its direction of motion describes a circular path, the normal acceleration being v^2/r where r is the radius of the path. Hence the centrifugal force on the body is mv^2/r where m is the body's mass. The equation of motion of the electron is, therefore,

$$\frac{mv^2}{r} = Hev \quad . \quad . \quad . \quad . \quad (3\text{–}3)$$

or

$$\frac{mv}{e} = Hr \quad . \quad . \quad . \quad . \quad (3\text{–}4)$$

If H and r can be determined, then the right-hand member of (3–4) may be found. The value of r is obtained by means of a discharge tube arranged in the manner indicated by Fig. 3–1. Cathode rays leave the cathode and are concentrated into a narrow pencil by means of a disc possessing a small aperture through which a proportion of the rays pass. The pencil is then deflected by passing through a magnetic field which is perpendicular to the plane of the paper. At the extremity of the tube is a screen coated with material which fluoresces under the impact of the cathode-ray pencil. By noting the displacement of the fluorescent spot when the field is applied, r may be determined. The exact method of doing this will be given later.*

If the electron traverses an electrostatic field, it will experience a force eX, where X is the electrostatic field strength. By arranging this field to be perpendicular to the magnetic field (i.e. parallel to the plane of the paper) and to occupy the same part of the path, the field strengths may be so adjusted that no deflexion of the fluorescent spot results. In such circumstances

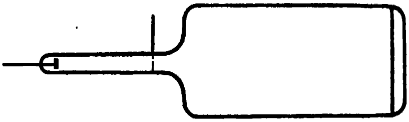


FIG. 3–1

$$eX = Hev$$

and
$$v = \frac{X}{H}$$

Thus, from a knowledge of the field strengths the electron velocity may be found. From the first experiment mv/e is obtained and hence from both experiments the ratio m/e is known.

By the above methods the following results were obtained by Sir J. J. Thomson—

GAS	v	m/e
Air	$2{\cdot}8 \times 10^9$	$1{\cdot}3 \times 10^{-7}$
,,	$2{\cdot}8 \times 10^9$	$1{\cdot}1 \times 10^{-7}$
,,	$2{\cdot}3 \times 10^9$	$1{\cdot}2 \times 10^{-7}$
,,	$2{\cdot}8 \times 10^9$	$1{\cdot}1 \times 10^{-7}$
Hydrogen	$2{\cdot}5 \times 10^9$	$1{\cdot}5 \times 10^{-7}$
CO	$2{\cdot}2 \times 10^9$	$1{\cdot}5 \times 10^{-7}$

---

*See p. 123.

It will be noted that the electron velocity is of a very high order, being about one-tenth of the velocity of light. It is also evident that the ratio m/e is independent of any particular gas filling.

KAUFMANN'S DETERMINATION OF e/m

The apparatus employed by Kaufmann in his determination of e/m is shown by Fig. 3–2. The discharge tube is furnished with an anode in the form of a thin wire, of which the cathode rays form a shadow on the plate P, which closes the end of the tube. The potential difference between C and A is measured with an electrostatic voltmeter. Surrounding the tube is a coil which produces a practically uniform field perpendicular to the cathode rays. Due to the field the cathode rays are deflected and the shadow of the anode takes up a different position on the plate P. As in Thomson's method, the electrons tend to follow a circular path under the influence of the field, and from (3–4) we have

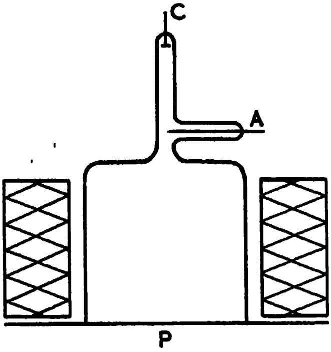


FIG. 3–2

$$\frac{mv^2}{r} = Hev \qquad (3\text{–}5)$$

Also if V is the cathode-anode potential difference

$$\tfrac{1}{2}mv^2 = eV \qquad (3\text{–}6)$$

From (3–5) and (3–6)

$$\frac{e}{m} = \frac{2V}{H^2r^2}$$

and

$$v = \frac{2V}{Hr}$$

r is determined from the deflexion of the anode shadow and the dimensions of the apparatus.

In Kaufmann's experiments V was varied from 3000 to 10,000 volts by varying the gas pressure in the discharge tube from 0·07 mm. to 0·03 mm. The final result obtained was $e/m = 1{\cdot}77 \times 10^7$ e.m.u. per gm., a result which differs very little from the now generally accepted value of $1{\cdot}757 \times 10^7$ e.m.u.

TOLMAN AND STEWART'S DETERMINATION OF e/m

An ingenious and quite different method of determining the ratio e/m was devised by Tolman and Stewart, a method which, in addition, tends to prove that electrical conduction in metals takes place by means of electrons. (See Chapter IV.) The principle of the method is that if "free" electrons exist within a conductor, then should a conductor be suddenly stopped while travelling at high speed, the inertia of the electrons will set up an e.m.f. Now, if a potential difference V is applied to the ends of a conductor of length l, the field strength is $X = V/l$. We have

$$ma = Xe = Ve/l$$

and

$$a = \frac{Ve}{lm}$$

where a is the acceleration imparted to an electron by the field X. Conversely, if a wire of length l is quickly brought to rest, the electrons suffer a negative acceleration a. This produces a potential difference between the ends of the wire given by

$$V = \frac{lma}{e}$$

The electrons, by reason of their inertia, tend to keep moving and this movement constitutes a current i. Thus

$$i = \frac{V}{R} = \frac{lma}{Re}$$

where R is the resistance of the wire. In Tolman and Stewart's experiment a flat coil of wire, comprised of many turns, was rotated at a high speed about an axis perpendicular to the plane of the coil and then suddenly stopped. Wires were taken from the coil via slip rings to a ballistic galvanometer. If t is the time the coil takes to come to rest, then the charge q passing through the galvanometer is it, so that

$$q = \frac{lmat}{Re} = \frac{lmv}{Re}$$

where v is the original velocity of the coil. Thus

$$\frac{e}{m} = \frac{lv}{Rq}$$

the value found for copper being 1·6 × 10^7 e.m.u./gm. The direction of the throw of the galvanometer was found to be that required by negatively-charged particles.

In another experiment a metal disc was spun at high speed, this causing electrons to be driven towards the periphery of the disc by centrifugal action. Thus, electrons collect at the edge of the disc, making this negative to the centre, until the field they produce just balances the effect of centrifugal force. The force on an electron due to the field is Xe, and that due to centrifugal action mv^2/r, where v and r are, respectively, the velocity and radius of the disc. Hence

$$Xe = mv^2/r$$

or

$$\frac{Ve}{r} = \frac{mv^2}{r}$$

and

$$V = \frac{mv^2}{e}$$

where V is the potential difference between the centre and edge of the disc. From the last expression

$$\frac{e}{m} = \frac{v^2}{V}$$

which was found to be the value for the electron.

So far, only the ratio e/m has been obtained and to find the values of m and e either one or the other must be directly determined.

The Wilson Cloud Chamber

In order to measure the charge of the electron, Thomson employed a Wilson cloud chamber. The cloud chamber is due to a discovery of Aitken that when condensation of water vapour occurs, it tends to do so on nuclei of dust or particles in the atmosphere rather than on air molecules. The method of operating a cloud chamber as employed by Aitken may be appreciated from Fig. 3–3. The globe G is initially filled with ordinary air and a small quantity of water to ensure complete saturation of the air. The taps t_1, t_2 being turned off, the pump exhausts R. t_2 is now turned on, with the result that an adiabatic expansion occurs in G, and the air cools and becomes supersaturated at the lowered temperature. The excess of moisture condenses in the form of fog or mist, the cloud slowly falling on to the floor of the cloud chamber G. If now a fresh supply of air is admitted to G through the filter, there will be fewer dust nuclei present. Hence, if a fresh expansion is made, the drops will be fewer and larger and the mist less dense. If the process is repeated several times the drops become easily visible and move downwards at an appreciable speed as a shower of rain. Ultimately, when the air is completely freed from dust, further expansions do not produce a mist or rain at all.

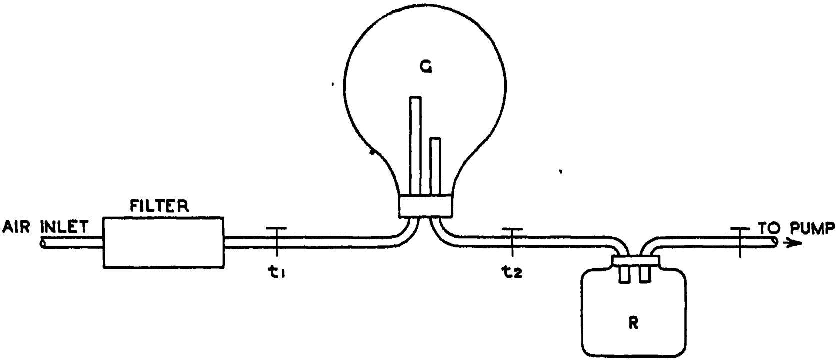


FIG. 3–3

An important aspect of the cloud chamber, as we shall presently see, is its ability to form clouds or fog tracks on ions and electrons.

The action of nuclei in condensing vapour is due to the principle that the vapour pressure over a surface depends on the curvature of that surface. For example, the vapour pressure over a convex surface is greater than over a plane one. When water evaporates from a plane surface there is no reduction in area and consequently no change in the potential energy due to surface tension. For the case of a curved surface, such as a water drop, evaporation will cause a reduction in the surface area and therefore a reduction in the potential energy due to surface tension. Hence surface tension will promote evaporation in this case which will go on farther from a spherical drop, than from a plane surface; i.e. the vapour pressure will be greater than over a plane surface.

To further these considerations, suppose a capillary tube dips in a liquid contained in a closed vessel (Fig. 3–4) and that no gas other than the vapour of the liquid is present above the liquid surface. Let p_1 and p_2 be the vapour pressures of the liquid at A and B respectively. If h is the height of B above A, then

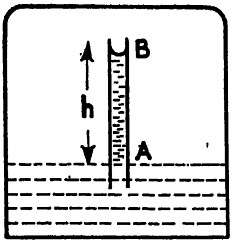


Fig. 3–4

$$p_1 = p_2 + g\sigma h \quad . \quad . \quad (3\text{–}7)$$

where σ is the vapour density, assumed constant as h is small. We may express h in terms of the surface tension T and the radius of curvature of the liquid at B. The pressure in the liquid just under the surface is

$$p_2 - \frac{2T}{r}.$$

and the liquid pressure at A is therefore

$$p_2 - \frac{2T}{r} + g\rho h$$

where ρ is the density of the liquid. But this is equal to p_1 and hence

$$p_1 = p_2 - \frac{2T}{r} + g\rho h$$

and, from (3–7)

$$p_2 + g\sigma h = p_2 - \frac{2T}{r} + g\rho h$$

or

$$\frac{2T}{r} = gh(\rho - \sigma)$$

and

$$p_1 - p_2 = gh\sigma = \frac{2T}{r}\left(\frac{\sigma}{\rho - \sigma}\right)$$

If the surface at B is convex, then

$$p_2 - p_1 = \frac{2T}{r}\left(\frac{\sigma}{\rho - \sigma}\right)$$

It will be noted that the vapour pressure over a curved surface differs from that over a plane surface by an amount

$$\frac{2T}{r}\left(\frac{\sigma}{\rho - \sigma}\right)$$

Unless the drops are very small the effect of curvature is inappreciable in altering the vapour pressure. When a drop is extremely small it will rapidly evaporate in an atmosphere from which vapour would condense on a plane surface. Hence, water drops which have to grow from indefinitely small drops by precipitation will only do so if nuclei such as dust particles are present or the atmosphere is supersaturated.

CONDENSATION AND FORMATION OF DROPS ON IONS

C. T. R. Wilson discovered that ions and electrons may form nuclei on which drops may grow by condensation. With dust-free air no cloud is formed if the expansion ratio is less than 1·25. If X-radiation is directed on the cloud chamber an expansion of 1·25 gives a cloud on electrons, but 1·34 is required for a cloud on positive ions. The following explanation may show the means whereby drops form on ions. The work done, per unit surface area, in increasing the radius of a drop by dr, is

$$\frac{2T}{r}\,dr$$

Thus the total work done due to surface tension is

$$4\pi r^2 \frac{2T}{r}\,dr = 8\pi rT dr$$

and the potential energy of the drop at radius r is

$$8\pi T \int_0^r r dr = 4\pi r^2 T$$

If the drop has a charge e, the potential energy due to the charge is $e^2/2Kr$, where K is the specific inductive capacity of the medium in which the drop exists. Thus, the total energy is

$$E = 4\pi r^2 T + \frac{e^2}{2Kr}$$

This energy will tend to become a minimum

i.e. $$\frac{dE}{dr} = 8\pi rT - \frac{e^2}{2Kr^2} = 0$$

or $$r^3 = \frac{e^2}{16\pi KT}$$

If a drop commences to form on an ion, the radius of which is, of course, exceedingly small, the drop will commence to grow and condensation will occur. This will go on until $r^3 = e^2/16\pi KT$. If

r^3 tends to exceed this value, the effect of surface tension will be to cause evaporation.

The Determination of *e*, The Electronic Charge

THOMSON'S METHOD

The apparatus employed by Thomson is illustrated by Fig. 3–5. The gas under experiment is situated in *A* and is ionized by X-radiation from above. The gas in the expansion chamber is shut in by a movable piston *P*, consisting of the end of a test tube. The joint is made air-tight with water, which also serves to keep the space *A* saturated with water vapour. The air space within *P* communicates with *B* via *T*, and the piston can be raised to any desired height by forcing air in at *C*. The vessel *X* is maintained at a low pressure and is closed by a valve *V*. On pulling back the valve rod *R*, the space under *P* is rapidly exhausted, and the piston falls, producing an adiabatic expansion of the gas above *P*. Let π be the atmospheric pressure, p_1 the pressure difference of the two arms of an attached manometer, before expansion, and σ the water vapour pressure. Then

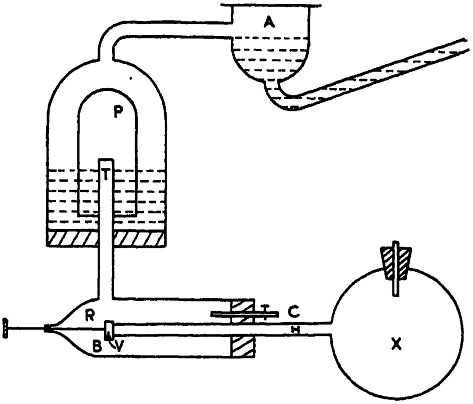


FIG. 3–5

$$\pi = P_1 + p_1 + \sigma \quad . \quad . \quad . \quad (3\text{–}8)$$

where P_1 is the gas pressure in *A* before expansion. After expansion and when the temperature of the apparatus has attained the temperature of the laboratory

$$P_2 = \pi - p_2 - \sigma \quad . \quad . \quad . \quad (3\text{–}9)$$

From Boyle's Law, $P_1 v_1 = P_2 v_2$, and from (3–8) and (3–9)

$$\frac{v_2}{v_1} = \frac{\pi - p_1 - \sigma}{\pi - p_2 - \sigma}$$

which gives the expansion. Also

$$v_1^{\gamma - 1}\,\theta_1 = v_2^{\gamma - 1}\,\theta_2$$

and

$$\log \frac{\theta_1}{\theta_2} = (\gamma - 1) \log \frac{v_2}{v_1}$$

which gives θ_2, the lowest temperature reached during the expansion.

Now, let W_1 be the weight of water vapour needed to saturate the expansion chamber at the initial temperature θ_1 and W_2 that to saturate it at θ_2. Then the weight of water deposited is $M = W_1 - W_2$. The drops constituting the cloud fall under gravity with a limiting velocity given by Stoke's Law, $s = \frac{2}{9}\frac{g\rho r^2}{\eta}$ where ρ is the density of water, r the radius of the drop, and η the viscosity of air. The velocity is found by observing the rate at which the surface of the cloud falls. Knowing this, r is found. If n is the number of drops in the cloud

$$\tfrac{4}{3}\,\pi r^3 \rho n = M$$

which determines n. The water surface in A is connected to an electrometer and thus the charge Q brought down by the cloud is obtained. Finally the charge on an ion or electron is Q/n.

The value of e obtained by Thomson by the foregoing method was $3{\cdot}4 \times 10^{-10}$ electrostatic units, a result now known to be somewhat low. The weak points in the method are the assumptions that the drops each contain but one ion and that the total mass of water in the cloud is that due to an adiabatic expansion. Actually the cloud tends to evaporate as the temperature of the air rises, with the result that the mass of the drops is not constant. Nevertheless, Thomson's investigation is justly celebrated and may be regarded as the first direct determination of the electronic charge.

H. A. Wilson's Method

A method of determining e which does not necessitate the determination of the mass of the cloud and the number of drops is due to H. A. Wilson. The velocity of a sphere moving through a viscous liquid is proportional to the force driving it, so that a vertical field X which exerts a force Xe on the drops will modify their velocity. If v_1 is rate of fall with $X = 0$, and v_2 the rate of fall in the field, then,

$$\frac{v_1}{v_2} = \frac{mg}{mg + Xe}$$

where m is the mass of a drop and the force Xe is reckoned positive when its direction is downwards.

The apparatus employed consisted of an expansion chamber containing two horizontal parallel electrodes between which the field X was maintained by means of a potential difference of about 2000 volts. The electrodes were $3{\cdot}5$ cm. in diameter and about

5 mm. apart. It was found that without the field the drops formed by an expansion all fell at the same rate. With the field applied this was not so, several sets of drops being discernible, each set falling at the same rate. Three such sets could usually be detected, it being found that the charges carried by them were in the ratio of 1 : 2 : 3. This shows that the charges are multiples of an atomic unit, giving confirmation of the theory of atomicity of electricity. The results obtained for e varied from 2×10^{-10} to $4{\cdot}4 \times 10^{-10}$ electrostatic units with a mean value of $3{\cdot}1 \times 10^{-10}$.

As with Thomson's method, inaccuracy was present due to evaporation of the drops.

MILLIKAN'S METHOD

Probably the most accurate determination of the electronic charge is due to Millikan, who employed an improvement of Wilson's method. Drops of oil or mercury were employed, which do not evaporate, these being produced by a sprayer or atomizer. The final form of Millikan's apparatus is shown by Fig. 3–6. *A* and *B* are similar circular metal plates, 22 cm. in diameter, 14·9 cm. apart, which have optically worked plane parallel surfaces. In the centre of the upper plate is a small hole through which drops of oil or mercury may pass from the sprayer above. The plates are supported in a metal box which is immersed in an oil tank *TT*. The drops are illuminated by means of an arc lamp through a window *W*, and are observed by means of a long-focus microscope, in which the image of a drop appears as a bright point of light.

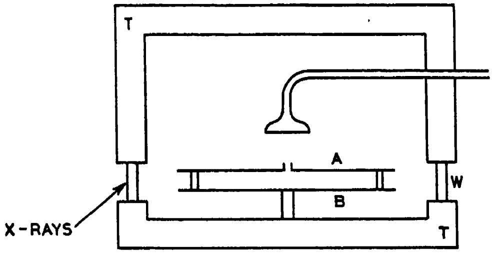


FIG. 3–6

On operating the sprayer, a large number of drops is formed above the plates, a few of which pass through the hole in the upper plate. X-radiation is passed between the plates, whereupon the drops pick up charges and, due to the electric field, either remain suspended or move slowly up or down. A drop may be selected and its velocity found by measuring the time its image in the microscope takes to pass from one hair line to another in the eyepiece. As with Wilson's method, we have

$$\frac{v_1}{v_2} = \frac{mg}{mg \pm Xe}$$

and $$e = \pm \frac{v_2 - v_1}{v_1} \cdot \frac{mg}{X}$$

m is determined from v_1 and Stoke's Law, and we have

$$m = \frac{4\pi\sigma}{3}\left(\frac{9}{2g\sigma}\right)^{\frac{3}{2}} v_1^{\frac{3}{2}}$$

which gives $$e = \pm \frac{4\pi}{3}\left(\frac{9}{2}\right)^{\frac{3}{2}}\left(\frac{1}{g\sigma}\right)^{\frac{1}{2}} \frac{(v_2 - v_1)v^{\frac{1}{2}}}{X} \quad . \quad . \quad (3\text{–}10)$$

From (3–10) it will be noted that the charge on a drop is proportional to $(v_2 - v_1)$. It was found that the values of $(v_2 - v_1)$ were always multiples of some minimum value $(v_2 - v_1)_0$. This shows that the charges on a drop were always multiples of a definite unit, i.e. the electronic charge, the charge being varied by the addition of electrons or positive ions one at a time. The final result obtained by Millikan was

$$e = 4{\cdot}774 \times 10^{-10} \text{ electrostatic units}$$

The value of e now generally accepted is $4{\cdot}77 \times 10^{-10}$ e.s.u. or $1{\cdot}59 \times 10^{-20}$ e.m.u. Combining this with the accepted value of $1{\cdot}757 \times 10^7$ for e/m gives $m = 9{\cdot}1 \times 10^{-28}$ gm. for the mass of the electron. Hence the mass of an electron is 1/1839 of that of an atom of the lightest known element, i.e. hydrogen.

ELECTRON DIMENSIONS

Having determined the charge and mass of the electron, it is evidently important to have some idea of its dimensions. As will be shown later, it is probable that an electron does not possess what is commonly termed inertia or gravitational mass, but owes its mass to the electromagnetic field it creates when set in motion. In a state of rest an electron possesses a radial electric field only. When in motion it behaves as a current element ev with an accompanying magnetic field. The energy of the latter is $H^2/8\pi$ ergs per unit volume, and the total energy may be equated to $\frac{1}{2}mv^2$, where v is the electron velocity. This means that it is assumed that the kinetic energy of the electron is entirely due to the electromagnetic field. The field will extend throughout space and, if it be assumed that the electron is spherical, will have the radius of the electron as a lower limit. Thus the charge of an electron is assumed to reside on, or within, a sphere of radius a. Now, according to an original assumption of Laplace, the field due to a current element ids is $ids\ \sin\theta/r^2$ where r is the distance from the element and θ is the angle between r and ds. (See Fig. 3–7.) Hence the field due to a moving electron is

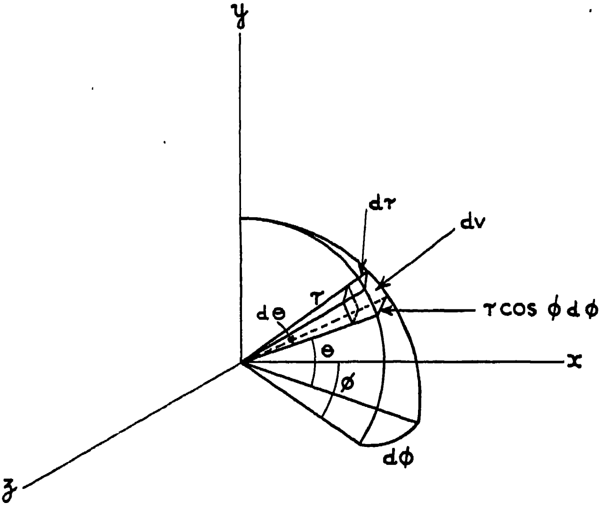


FIG. 3–7

$$H = \frac{ev \sin \theta}{r^2}$$

and the energy per unit volume at any position in space is

$$E = \frac{(ev \sin\theta/r^2)^2}{8\pi}$$

Taking a volume element dV, the total energy of the field is given by the integral

$$E_t = \int E dV$$

where this is taken over the whole of space surrounding the electron. Referring to Fig. 3–7, the volume-element dV may be expressed in spherical polar co-ordinates with the electron situated at the origin and having its velocity directed along the x axis. Thus, from Fig. 3–7

$$dV = rd\theta \times dr \times r\cos\theta d\phi$$

and
$$E_t = \int \frac{e^2v^2}{8\pi r^4}\sin^2\theta r^2 dr d\phi \cos\theta d\theta$$

Expressing this as a triple integral with appropriate limits

$$E_t = \frac{e^2v^2}{8\pi}\int_a^\infty \frac{dr}{r^2}\int_0^{2\pi} d\phi \int_{-\pi/2}^{\pi/2} \sin^2\theta \cos\theta d\theta$$

$$= \frac{e^2v^2}{3a}$$

But $E_t = \frac{1}{2}mv^2$, and thus

$$\frac{e^2v^2}{3a} = \frac{1}{2}mv^2$$

which gives
$$a = \frac{2}{3}\frac{e}{m}$$

Substituting for e and m, $a = 1{\cdot}9 \times 10^{-13}$ cm. It may be mentioned that, as we are dealing with an electromagnetic field, e must be expressed in e.m.u.

Electron Dynamics

It will have been noted that the majority of experiments devised to study the properties of ions and electrons depend on the behaviour of the latter under the action of magnetic or electric fields or combinations of both. Furthermore, scientific and engineering applications of ions and electrons (i.e. Electronics) in the majority of instances depend for their functioning on the motion of these particles under the influence of such fields. This motion may be termed electron dynamics, and to this subject closer attention will now be directed. It may be found that some of the ground already covered in previous chapters will be touched on, but some of this matter will be included here for the sake of completeness. The simplest examples will be first treated, advancing to others of greater complexity.

MOTION OF A CHARGED PARTICLE IN AN ELECTRIC FIELD

The force on a charge q in an electric field of strength X is given by

$$qX \quad . \quad . \quad . \quad . \quad (3\text{–}11)$$

and in the absence of impeding forces the charge receives an acceleration equal to

$$\frac{q}{m} X \quad . \quad . \quad . \quad . \quad (3\text{–}12)$$

where m is the mass of the particle. The charge, of course, moves in the direction of the force, and after a time t possesses a velocity

$$u = \frac{q}{m} Xt + u_0 \quad . \quad . \quad . \quad (3\text{–}13)$$

where u_0 is the initial velocity. Assuming the direction of flight is parallel to the x axis, the distance of the particle from the origin at time t is

$$x = \frac{1}{2}\frac{q}{m} Xt^2 + u_0 t + x_0 \quad . \quad . \quad . \quad (3\text{–}14)$$

where x_0 is the position from the origin when $t = 0$. The velocity of the charge may be expressed in terms of the potential difference through which it has fallen. Thus

$$\tfrac{1}{2}m(u^2 - u_0{}^2) = q(V - V_0) \quad . \quad . \quad . \quad (3\text{–}15)$$

where V_0 is the potential at the position corresponding to u_0 and V similarly corresponds to u. From (3–15)

$$(u^2 - u_0{}^2) = \frac{2q}{m}(V - V_0)$$

and if V_0 and $u_0 = 0$

$$u = \sqrt{\frac{2q}{m}}\, V^{\frac{1}{2}} = \frac{q}{m} Xt \quad . \quad . \quad . \quad (3\text{–}16)$$

Substituting from (3–16) into (3–14)

$$x = \tfrac{1}{2}\sqrt{\frac{2q}{m}}\, V^{\frac{1}{2}}t + u_0 t + x_0$$

or, if u_0 and $x_0 = 0$,

$$t = \sqrt{\frac{2m}{q}}\, V^{-\frac{1}{2}}x \quad . \quad . \quad . \quad . \quad (3\text{–}17)$$

which gives the time of flight. When q and m are, respectively, the electronic charge and mass, and V is expressed in volts

$$\sqrt{\frac{2e}{m}} = 5{\cdot}95 \times 10^7 \text{ cm. per volt}^{\frac{1}{2}}\text{/sec.}$$

For an ion of atomic weight A

$$\sqrt{\frac{2e}{m}} = 1{\cdot}39 \times 10^6 A^{-\frac{1}{2}}\, 10' \text{cm. per volt}^{\frac{1}{2}}\text{/sec}$$

The foregoing equations are, of course, based on the assumption that X is constant. If X is a function of time, then

$$\frac{du}{dt} = \frac{q}{m}X$$

and
$$u = \frac{q}{m}\int X dt$$

Also
$$\frac{dx}{dt} = u + u_0$$

and
$$x = \int (u + u_0)dt$$

The expression given by (3–15) is an energy equation and depends only on the initial and final states. Hence, if $u_0 = 0$, u is always equal to $\sqrt{2q/m}\,V^{\frac{1}{2}}$ and is independent on the manner in which V varies with respect to time. The time of flight of an electron, or charged particle, must, however, be expressed differently from (3–16). We have

$$\frac{dx}{dt} = u = \sqrt{\frac{2q}{m}}\,V^{\frac{1}{2}}$$

$$dt = \sqrt{\frac{m}{2q}}\,V^{-\frac{1}{2}}dx$$

and
$$t = \sqrt{\frac{m}{2q}}\int V^{-\frac{1}{2}}dx$$

If X is constant, then $V = Xx$ and

$$\int V^{-\frac{1}{2}}dx = \int (Xx)^{-\frac{1}{2}}dx = 2(Xx)^{-\frac{1}{2}}x$$

and
$$t = \sqrt{\frac{2m}{q}}\,V^{-\frac{1}{2}}x \text{ as before.}$$

MOTION IN TWO DIMENSIONS

A case of considerable practical importance is where the initial velocity direction of an electron does not coincide with that of the field. Referring to Fig. 3–8 (a), let the field be in the direction of the y-axis and the electron enter the field at an angle given by

$$\tan\theta_0 = v_0/u_0$$

where v_0 and u_0 are, respectively, the initial velocity components of the electron in the y and x directions. As the x-velocity component remains unaffected, we have

$$u = u_0$$

$$x = u_0 t$$

t being reckoned from the moment the electron enters the field. Again

$$v = \frac{q}{m} Xt + v_0$$

$$y = \frac{1}{2}\frac{q}{m} Xt^2 + v_0 t$$

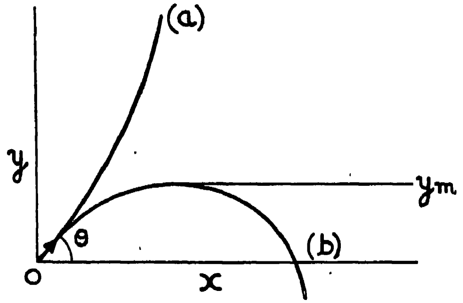


FIG. 3–8

Assuming X to be independent of t. Eliminating t we have

$$v = \frac{q}{m} X \frac{x}{u_0} + v_0 \quad . \quad . \quad . \quad . \quad (3\text{–}18)$$

$$y = \frac{1}{2}\frac{q}{m} X \frac{x^2}{u_0^2} + \frac{v_0}{u_0} x \quad . \quad . \quad . \quad (3\text{–}19)$$

which show that the v-directed velocity component is a linear function of x and that the path of the electron is a parabola. The angle of flight is given by

$$\tan \theta = \frac{v}{u} = \frac{q}{m} X \frac{x}{u_0^2} + \frac{v_0}{u_0}$$

The foregoing equations for two-dimensional motion are based on the assumption that the field vector points in the -y direction. If the opposite direction is assumed the electron trajectory will have a maximum for a finite value of x. In this case we must write

$$v = -\frac{q}{m}Xt + v_0$$

$$y = -\frac{1}{2}\frac{q}{m}Xt^2 + v_0 t$$

or

$$v = -\frac{q}{m}X\frac{x}{u_0} + v_0 \quad . \quad . \quad . \quad (3\text{–}20)$$

$$y = -\frac{1}{2}\frac{q}{m}X\frac{x^2}{u_0^2} + \frac{v_0}{u_0}x \quad . \quad . \quad . \quad (3\text{–}21)$$

For a maximum

$$\frac{dy}{dx} = -\frac{q}{m}X\frac{x}{u_0^2} + \frac{v_0}{u_0} = 0$$

or

$$x = \frac{m}{q}\frac{v_0 u_0}{X}$$

and

$$y_m = \frac{1}{2}\frac{m}{q}\frac{v_0^2}{X} \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}22)$$

A case of common occurrence is where $v_0 = 0$. In this circumstance, the trajectories are identical in form (although opposite in direction) whether the direction of the field is positive or negative.

VARIATION OF MASS WITH VELOCITY

In the treatment of the motion of a charged particle it has so far been assumed that the mass of the particle is unaffected by its velocity. This assumption is justifiable, provided that the velocity is less than that corresponding to about 3000 electron-volts, but above this it may be necessary to introduce a correction factor. To those familiar with the Special Theory of Relativity it is, of course, well known that the mass of a body is related to its velocity by the following equation

$$m = \frac{m_0}{\sqrt{1 - (v/c)^2}}$$

where m_0 is the rest-mass of the body, v its velocity, and c the velocity of light. It is, however, possible to deduce this result from electromagnetic theory and this procedure will now be followed.

ELECTROMAGNETIC MOMENTUM

According to Maxwell's electromagnetic theory of radiation, an electromagnetic wave exerts a pressure on the surface on which it falls. If total absorption occurs this pressure is equal to the energy of the wave per unit volume. Let E be the pressure (and energy) per unit volume of the wave, and consider a volume of area A and length x moving with velocity c. If this energy is assumed to possess momentum and M is its magnitude, then, from Newton's Second Law of Motion

$$f = \frac{dM}{dt} = EA \quad . \quad . \quad . \quad (3\text{–}23)$$

because EA is the force exerted when total absorption occurs. Let dx be the distance travelled by the wave in time dt. Multiplying both sides of (3–23) by dx we have

$$\frac{dM}{dt}\,dx = EAdx$$

or

$$dM\,\frac{dx}{dt} = EAdx$$

But $dx/dt = c$; therefore

$$dM = \frac{EAdx}{c}$$

Integrating

$$\int dM = \int_0^x \frac{EAdx}{c}$$

or

$$M = \frac{E}{c}\,.\,Ax$$

As Ax is the volume, it follows that the momentum of the wave per unit volume is

$$M = \frac{E}{c} \quad . \quad . \quad . \quad . \quad (3\text{–}24)$$

From this result it appears that a field has energy and momentum, and as it can travel through space it has some of the essential properties of matter. As this momentum is that of the energy of the field it may be said that electromagnetic energy in motion has momentum. It may be supposed, therefore, that the momentum of matter is simply the momentum of the energy it contains. This follows because energy can be converted from one form to another, and if electromagnetic energy has momentum it is probable that other forms of energy have the same amount and kind of momentum.

ELECTROMAGNETIC MASS

As a field possesses energy and momentum, two essential properties of matter, it is possible that it possesses a third essential property, i.e. mass. This mass may be determined in the following manner. Referring to Fig. 3–9, let this represent a radiation-tight tube with bodies A and B of equal mass at its ends. Assume that A has initially more energy than B and that it discharges this excess energy in the form of an electromagnetic wave to B. As the wave leaves, A experiences a backward impulse I, and B experiences a forward impulse I when it arrives, due to the wave possessing momentum. Let d be the distance from A to B. Then during the time d/c the system travels towards the left with velocity I/m_{AB} where m_{AB} is the mass of the bodies A and B. The distance the system is displaced during this time is

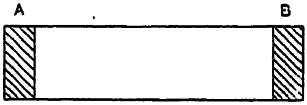


FIG. 3–9

$$\frac{I}{m_{AB}} \cdot \frac{d}{c}$$

but by Newton's Third Law the centroid remains in the same place. Thus, if the mass m_{AB} is displaced to the left, another mass must be displaced to the right. The only thing which moves to the right is the wave which travels the distance d. Hence if the mass of the wave is m

$$md = m_{AB} \frac{I}{m_{AB}} \cdot \frac{d}{c}$$

or

$$m = I/c$$

But $I = E_t/c$, and hence

$$m = \frac{E_t}{c^2} \quad . \quad . \quad . \quad . \quad . \quad (3–25)$$

where E_t is the total energy of the wave. From this it will be seen that the mass of the field per unit volume is E/c^2.

From the foregoing it appears that electromagnetic energy may be considered to possess mass, and as this energy may be converted to other forms it is probable that all forms of energy have mass in accordance with the relation expressed by (3–25). It follows as a corollary that all mass may possess energy in accordance with (3–25).

RELATION BETWEEN ENERGY, MASS, AND VELOCITY OF A PARTICLE

Considering a particle of mass m moving with velocity v, its momentum may be written

$$M = \frac{Ev}{c^2}$$

If a force acts on it in the direction in which it is moving and there is no loss by radiation

$$f = \frac{dM}{dt} = \frac{E}{c^2}\frac{dv}{dt} + \frac{v}{c^2}\frac{dE}{dt}$$

Now $$fdt = dM$$

$$fdx = dE$$

$$dx = vdt$$

from which $$dM = dE/v$$

and $$\frac{dE}{v} = \frac{Edv + vdE}{c^2}$$

Transposing $$\frac{dE}{E} = \frac{vdv}{c^2 - v^2}$$

which leads to $$\log E = -\tfrac{1}{2}\log(c^2 - v^2) + \log C$$

When $v = 0$, $E = E_0$. Therefore

$$C = E_0 c$$

and $$E = \frac{E_0 c}{\sqrt{c^2 - v^2}}$$

$$= \frac{E_0}{\sqrt{1 - v^2/c^2}} \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}26)$$

From (3–25) we may write $E = mc^2$ and $E_0 = m_0c^2$, where m_0 is the rest-mass of the particle. Substituting in (3–26)

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}27)$$

a result which is identical with that derived from the theory of relativity. The kinetic energy of the particle is

$$E - E_0 = E_0\left[\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right] \quad . \quad . \quad . \quad (3\text{–}28)$$

If v/c is small (as it generally is), (3–28) becomes

$$E_0\left[1 + \frac{v^2}{2c^2} - 1\right]$$

$$= \frac{E_0 v^2}{2c^2}$$

and since $m_0 = E_0/c^2$, we have

$$E - E_0 = \tfrac{1}{2} m_0 v^2$$

It is evident from (3–28) that it is only when v is comparable with c that a correction must be applied for the variation in the mass of a body according to its velocity. If the velocity of a charged particle is calculated in the normal manner from the potential through which it has fallen, it is apparent that this calculation ignores the effect of the increased mass of the particle. Hence the calculated velocity is greater than the actual velocity. The latter can be found from the former in the following manner—

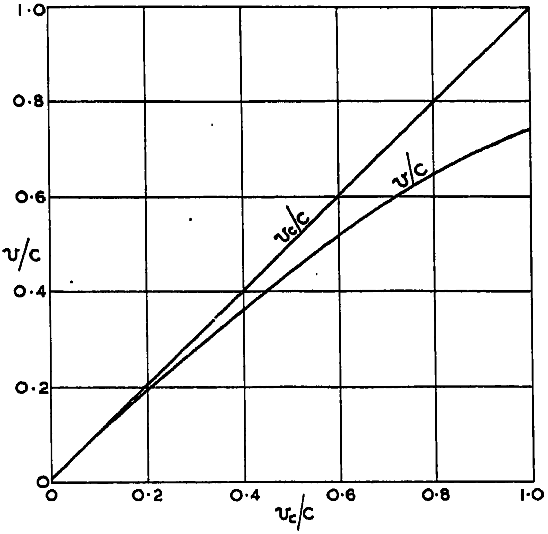


FIG. 3–10

The calculated velocity is given by

$$\tfrac{1}{2} m_0 v_c^2 = eV \qquad . \quad . \quad . \quad (3\text{-}29)$$

or

$$v_c^2 = \frac{2eV}{m_0}$$

Also from (3–28) and (3–29)

$$m_0c^2\left[\frac{1}{\sqrt{1-v^2/c^2}}-1\right]=eV=\tfrac{1}{2}m_0v_c^2$$

Hence $$\frac{v_c^2}{2c^2}+1=\frac{1}{\sqrt{1-v^2/c^2}}$$

and $$1-\frac{v^2}{c^2}=\frac{1}{\left(1+\dfrac{v_c^2}{2c^2}\right)^2}$$

which gives $$v=c\left[1-\frac{1}{\left(1+\dfrac{v_c^2}{2c^2}\right)^2}\right]^{\frac{1}{2}}$$

v being the true velocity and v_c that as calculated above. Referring to Fig. 3–10, it will be appreciated that it is only when v_c/c exceeds about 0·1 that it is necessary to replace v_c by v. At this value, $V = 2540$ electron-volts.

Electron Radius

With the assistance of the foregoing theory of electromagnetic mass the radius of the electron may be calculated. Assuming the charge of the electron to be distributed over a sphere of radius a, the energy of this charge is

$$= \tfrac{1}{2} \times \text{charge} \times \text{potential} = mc^2$$

or $$\tfrac{1}{2}e\,.\,\frac{e}{a}=mc^2$$

and $$a=\frac{e^2}{2mc^2}=\frac{(4{\cdot}774\times10^{-10})^2}{2\times9{\cdot}1\times10^{-28}\times(3\times10^{10})^2}\text{ cm.}$$

$$=1{\cdot}4\times10^{-13}\text{ cm.}$$

which is of the same order of magnitude as the result obtained on page 113.

Motion of a Charged Particle in a Magnetic Field

If a charged particle is placed within a magnetic field, then, provided there is no relative motion between field and particle, the latter will experience no force. It will be noted that this is a distinctly different state of affairs from that occurring with an electric field. If the particle is set in motion it will be equivalent to an element of current, for the latter is only a charge in motion.

Referring to Fig. 3–11, the force acting on this equivalent current element is

$$f = Hids \sin \theta$$

where H is the magnetic field strength, i the current in e.m.u., ds an element of length, and θ the angle between the directions of H and ds. Writing $i = dq/dt$ we have

$$\begin{aligned} f &= H \frac{dq}{dt} ds \sin \theta \\ &= Hdq \frac{ds}{dt} \sin \theta \\ &= Hqv \sin \theta \quad . \quad . \quad . \quad . \quad (3\text{–}30) \end{aligned}$$

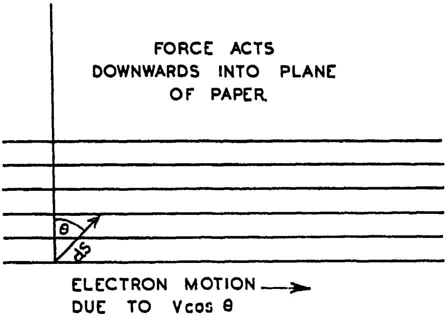

FORCE ACTS DOWNWARDS INTO PLANE OF PAPER
θ
ELECTRON MOTION ⟶ DUE TO Vcos θ

FIG. 3–11

where q and v are, respectively, the charge and velocity of the particle. This expression clearly shows the dependence of the force on the velocity of the particle and also indicates that when the motion is parallel to the field no force exists. The direction in which the force acts depends, of course, on the sign of q. Thus, the direction for an electron is opposite to that for a positive ion.

Let it now be assumed that an electron, or positive ion, with a velocity v enters a uniform magnetic field at an angle θ. The component of the velocity parallel to the field is $v \cos \theta$ and, as this does not produce a force, this component will be unaffected and the particle will move along the field with velocity $v \cos \theta$. The velocity component $v \sin \theta$ produces a constant force perpendicular to itself, and as this force has no component parallel to $v \sin \theta$ the *direction* of the particle will be changed but not its *speed*. Now a body moving in a circular path with constant speed experiences a constant force perpendicular to its direction of motion, the magnitude of this force being equal to mv^2/r, where m is the mass of the body and r the radius of its circular path. Hence it follows that the velocity component $v \sin \theta$ of a charged particle will cause it to describe a circular path in a constant magnetic field, the radius of the path being given by

$$Hqv \sin \theta = mv^2 \sin^2 \theta / r$$

i.e.

$$r = \frac{mv \sin \theta}{Hq} \text{ cm.} \qquad (3\text{–}31)$$

If the kinetic energy of the particle is expressed in electron-volts, we have

$$\tfrac{1}{2}mv^2 = qV$$

and

$$r = \frac{\sin \theta}{H}\sqrt{\frac{m}{q}}\,(2V)^{\frac{1}{2}}$$

which, for an electron, is

$$\frac{3{\cdot}36 \sin \theta}{H}\,(V)^{\frac{1}{2}} \qquad (3\text{–}32)$$

The angular velocity with which the particle describes its path is

$$w = \frac{v \sin \theta}{r} = \frac{Hq}{m}$$

which is independent of the linear velocity and θ.

The periodic time is

$$T = \frac{2\pi}{w} = \frac{2\pi m}{Hq} \qquad (3\text{–}33)$$

Now during the time the particle is traversing the circular path the velocity component $v \cos \theta$ is translating the particle parallel to the direction of the magnetic field. Hence, it follows that the actual path traversed by the particle is a helix, the pitch of which is

$$Tv \cos \theta = \frac{2\pi mv \cos \theta}{Hq} \qquad (3\text{–}34)$$

PARTICLE MOTION IN SIMULTANEOUS ELECTRIC AND MAGNETIC FIELDS

The motion of a charged particle under the simultaneous influence of electric and magnetic fields parallel to each other will first be considered. Referring to Fig. 3–12, the fields will be taken to be parallel to the y-axis with the directions shown. If the particle enters the fields without a velocity component in the x-direction it will simply move parallel to the field according to the relations

$$v = \frac{q}{m} Xt + v_0$$

$$y = \frac{1}{2}\frac{q}{m} Xt^2 + v_0 t \qquad . \quad . \quad . \quad (3\text{–}35)$$

As no velocity-component perpendicular to the magnetic field exists, this field will be without effect. Assuming now that a velocity-component, u_0, exists in the x-direction, this will give rise to circular motion, the radius of the path being independent of X.

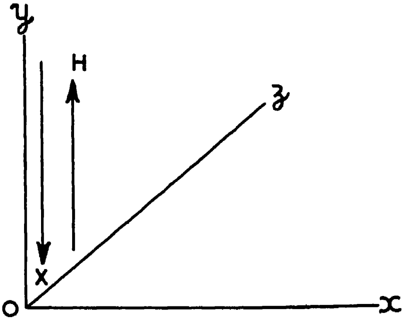


Fig. 3–12

The particle will now describe a helix along the y-axis, the pitch constantly increasing with time as indicated by (3–35). The value of the pitch is obtained from (3–33) and (3–35), and is

$$p = \frac{1}{2}\frac{q}{m} X[(t_1 + T)^2 - t_1^2] + v_0[(t_1 + T) - t_1]$$

$$= \frac{1}{2}\frac{q}{m} XT(2t_1 + T + v_0)$$

or

$$p = \pi \frac{X}{H}\left(2t_1 + \frac{2\pi m}{Hq} + v_0\right)$$

where t_1 is any instant of time. The radius of the circle is, of course, given by either (3–31) or (3–32).

Particle Motion in Mutually Perpendicular Electric and Magnetic Fields

In this case it will be assumed that the electric field is in the $-x$-direction and the magnetic field is in the $-z$-direction, as shown by Fig. 3–13. If the only initial velocity-component is v_0, then the force on the particle due to the magnetic field will be in the opposite direction to that due to the electric field. Hence the resultant force on the particle is

$$Hqv_0 - Xq$$

and, if $v_0 = X/H$, this is equal to zero. Thus, in this special case, the motion of the particle is unaltered by the joint action of the fields.

A case of some importance is when the particle is released with zero velocity, i.e. $u_0 = v_0 = w_0 = 0$. At any instant after release

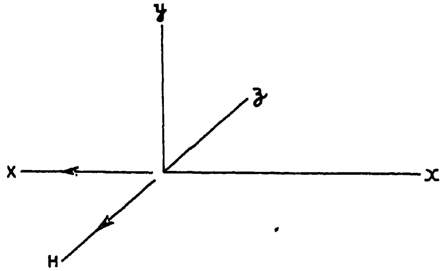


Fig. 3–13

the x velocity component is dx/dt. The force on the particle due to this component and the magnetic field is in the y-direction and is given by $Hq\,dx/dt$. Hence

$$m\frac{d^2y}{dt^2} = Hq\frac{dx}{dt} \qquad . \quad . \quad . \quad (3\text{–}36)$$

The force on the particle due to the electric field is Xq and acts in the x-direction. In addition to this force is one created by the magnetic field and the y-directed velocity component of the particle. This acts in the $-x$ direction, and hence the resultant force along the x axis is

$$m\frac{d^2x}{dt^2} = Xq - Hq\frac{dy}{dt} \qquad . \quad . \quad . \quad (3\text{–}37)$$

Differentiating (3–36) and (3–37)

$$m\frac{d^3y}{dt^3} = Hq\frac{d^2x}{dt^2} \qquad . \quad . \quad . \quad . \quad (3\text{–}38)$$

$$m\frac{d^3x}{dt^3} = -Hq\frac{d^2y}{dt^2} \qquad . \quad . \quad . \quad (3\text{–}39)$$

From (3–39) $$\frac{d^2y}{dt^2} = -\frac{m}{Hq}\frac{d^3x}{dt^3}$$

and substituting in (3–36) we have

$$-\frac{m^2}{Hq}\frac{d^3x}{dt^3} = Hq\frac{dx}{dt}$$

or $$m^2\frac{d^3x}{dt^3} + H^2q^2\frac{dx}{dt} = 0$$

Let $dx/dt = Z$. Then

$$m^2Z'' + H^2q^2Z = 0$$

and $$Z = A\sin\frac{Hq}{m}t + B\cos\frac{Hq}{m}t$$

$$x = -\frac{m}{Hq}A\cos\frac{Hq}{m}t + \frac{m}{Hq}B\sin\frac{Hq}{m}t + C$$

where A, B, and C are constants.

When $t = 0$, $x = 0$, $dx/dt = 0$.

Therefore $$B = 0 \text{ and } C = \frac{m}{Hq}A$$

and $$x = \frac{A}{p}(1 - \cos pt), \text{ where } p = \frac{Hq}{m}$$

In order to obtain A it may be noted that for $t = 0$, $dy/dt = 0$.

Referring to (3–37) we have

$$m\frac{d^2x}{dt^2} = Xq - Hq\frac{dy}{dt}$$

or $$mpA = Xq \text{ when } t = 0$$

Hence $$A = \frac{Xq}{mp} = \frac{X}{H}$$

and $$x = \frac{X}{pH}(1 - \cos pt) \quad . \quad . \quad . \quad (3\text{–}40)$$

In order to obtain y we differentiate (3–40) twice and substitute in (3–37), obtaining

$$\frac{mpX}{H}\cos pt = Xq - Hq\frac{dy}{dt}$$

$$\frac{dy}{dt} = \frac{X}{H} - \frac{mpX}{H^2q}\cos pt$$

and $$y = \frac{X}{H}t - \frac{mX}{H^2q}\sin pt + D$$

where D is a constant.

When $t = 0$, $y = 0$, therefore $D = 0$

and
$$y = \frac{X}{H}t - \frac{mX}{H^2q}\sin pt \qquad . \quad . \quad (3\text{–}41)$$

Collecting (3–40) and (3–41) we finally have

$$x = \frac{X}{pH}(1 - \cos pt) \quad . \quad . \quad . \quad (3\text{–}42)$$

$$y = \frac{X}{pH}(pt - \sin pt) \quad . \quad . \quad . \quad (3\text{–}43)$$

These last two equations are those of a cycloid, i.e. the curve generated by a point on the circumference of a circle of radius r which rolls along a straight line, the y-axis. Referring to Fig. 3–14, the position of the point P represents the position of the electron at any instant. Assuming this to have been initially at the origin, then its co-ordinates may be expressed in terms of θ, i.e. the angle through which the circle has rolled. The distance OA is equal to $r\theta$, this being equal to the length of the arc PA. The distance y is

$$r\theta - PC\sin\theta$$
$$= r(\theta - \sin\theta)$$

Also x is equal to
$$OC - PB$$
$$= r - r\cos\theta$$

Hence
$$x = r(1 - \cos\theta)$$
$$y = r(\theta - \sin\theta)$$

which is identical with (3–42) and (3–43) if $\theta = pt$ and $r = X/pH$.

It will be appreciated that the particle will describe a path similar to that of a point P on a rolling circle and that its distance from the y axis cannot exceed the diameter of the circle. Hence

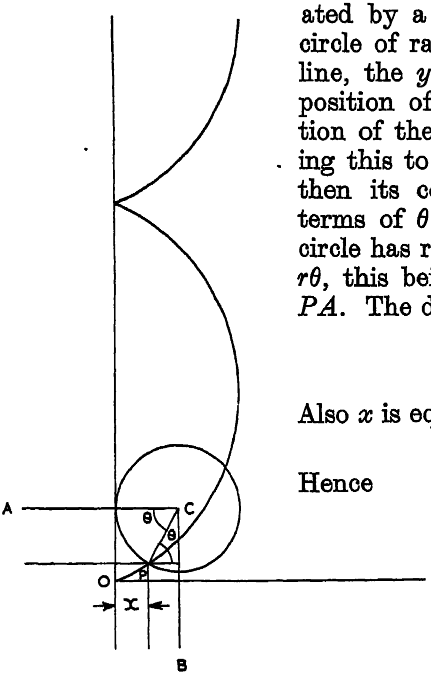


FIG. 3–14

$$2r = \frac{2X}{pH}$$

or
$$x_m = \frac{2X}{pH}$$

where x_m is the maximum distance the particle can travel in the x direction.

The cycloidal motion described has been employed to determine the ratio m/e for the electron. To effect this, a zinc plate takes the position of the y-axis and is exposed to ultra-violet light which causes it to emit electrons by photo-electric action. If a second plate is placed parallel to the zinc plate, no charge will be received by the former until the distance between the plates is

$$d = \frac{2X}{H^2} \cdot \frac{m}{e}$$

Thus m/e may be determined.

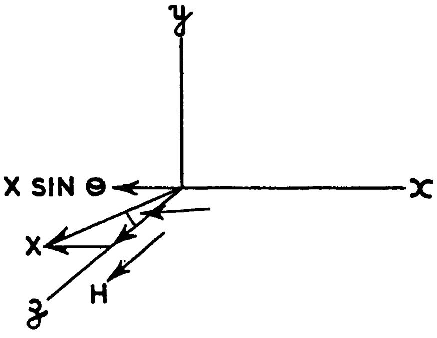


FIG. 3·15

The velocity of a charged particle undergoing cycloidal motion is given by

$$\sqrt{u^2 + v^2} = \sqrt{(dx/dt)^2 + (dy/dt)^2}$$

$$= \sqrt{\left(\frac{X}{H} \sin pt\right)^2 + \left[\frac{X}{H}(1 - \cos pt)\right]^2}$$

$$= \sqrt{\frac{2X^2}{H^2}(1 - \cos pt)}$$

or

$$\frac{2X}{H} \sin \frac{pt}{2}$$

Thus, when $pt = 2\pi$, i.e. after one complete period, the velocity of the particle is zero.

In addition to the various motions already considered, and referred to cartesian co-ordinates, we may consider one further and final case before passing to motions referred to a polar co-ordinate system. This case is where the directions of the electric and magnetic fields make an angle with each other. The plane determined by the fields will be taken as the xz plane, the conditions being shown by Fig. 3–15, where θ is the angle between the field directions. The electric field may be resolved into two components, one, $X \cos \theta$, parallel to the magnetic field and the other, $X \sin \theta$, perpendicular to the magnetic field and having the $-x$ direction. It follows that the conditions are similar to those of Fig. 3–13, with the exception that X in (3–42) and (3–43) must be replaced by $X \sin \theta$ and that now a z-co-ordinate exists due to $X \cos \theta$. Thus, the projection of the motion of the particle upon the xy plane is still a cycloid, but the motion now occurs in three dimensions instead of the former two. The co-ordinates as a function of t are

$$x = \frac{X \sin \theta}{pH} (1 - \cos pt)$$

$$y = \frac{X \sin \theta}{pH} (pt - \sin pt)$$

$$z = \frac{qX \cos \theta}{2m} t^2$$

it being assumed that the particle starts from rest.

CYLINDRICAL SYMMETRICAL SYSTEMS

The foregoing analyses of charged particle motion have been effected with the use of rectangular co-ordinates. However, in cases where cylindrical symmetry exists, such as, for example, between concentric cylinders, it is preferable to employ cylindrical polar co-ordinates. As an instance, consider a long straight filament surrounding which is a coaxial cylinder. If a potential difference is maintained between these, the cylinder being positive, a purely radial field will exist which will be cylindrically symmetrical. Should electrons be emitted from the filament, they will be accelerated by the field and travel radially to the cylinder. The electron acceleration is not constant because the field strength is a function of r, i.e. the radial distance from the filament axis. The value of X as a function of r may be found in the following manner. Let a charge Q exist per unit length of the filament. Then the charge density is $Q/2\pi r_f$ and the field at any radius r is $4\pi \,.\, Q/2\pi r = 2Q/r$. Hence

$$X = 2Q/r$$

The force on a charged particle at radius r is

$$qX = \frac{2qQ}{r}$$

and the work done in translating the particle a distance dr is

$$qXdr = \frac{2qQdr}{r}$$

If q is a unit charge, then the potential difference between the filament and cylinder is

$$V = 2Q\int_{r_f}^{r_c}\frac{dr}{r}$$

$$= 2Q \log \frac{r_c}{r_f}$$

Dividing both sides by r

$$\frac{V}{r} = \frac{2Q}{r} \log \frac{r_c}{r_f} = X \log \frac{r_c}{r_f}$$

and

$$X = \frac{V}{r} \cdot \frac{1}{\log \frac{r_c}{r_f}} \text{ e.s.u./cm.} \qquad . \quad . \quad . \quad (3\text{–}44)$$

From this result it will be noted that, unlike previously considered cases, the field is not constant, but is a function of r.

We shall now consider the case where a radial electric field exists simultaneously with a uniform magnetic field, the latter being parallel to the filament axis. Electrons emitted from the filament will possess radial and transverse motions due to these fields, and the motions will be considered with the assistance of the cylindrical co-ordinate system of Fig. 3–16. At any point P the force acting on an electron in the y direction is $f_y = ma_y$, where a_y is the y-directed acceleration. The y-directed velocity is, of course, dy/dt. The radial and transverse velocities are, respectively, dr/dt and $rd\theta/dt$. The radial acceleration may be considered to consist of two components, one due to the electron's purely radial motion, given by d^2r/dt^2, and the other due to the transverse motion, given by $-r\omega^2$ or $-r(d\theta/dt)^2$. The signs of these two components are opposite in character, because the first acts radially outwards and the other is always directed inwards towards the centre of circle of radius r. Hence the total radial acceleration is

$$a_r = [d^2r/dt^2 - r(d\theta/dt)^2]$$

Now the effect of the magnetic field is to produce a torque on th

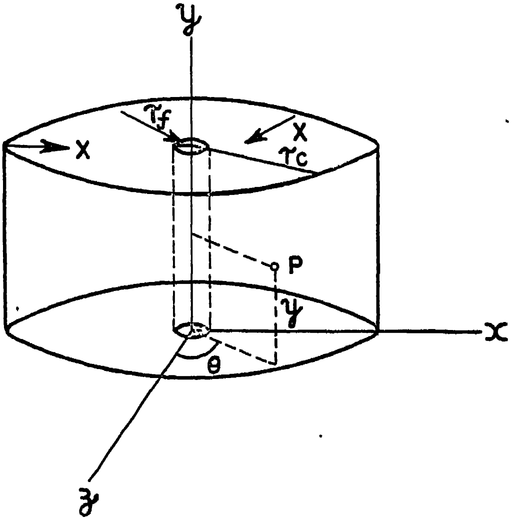


FIG. 3–16

electron at right-angles to its direction of motion. As torque equal to rate of change of angular momentum, we may write

$$T = f_\theta r = \frac{d(Iw)}{dt}$$

where f_0 is the tangential force, Iw the angular momentum, and the moment of inertia. Differentiating

$$f_0 r = \frac{d(mr^2w)}{dt}$$

$$= m\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right)$$

or $$m\left(r^2\frac{d^20}{dt^2} + 2r\frac{dr}{dt}\cdot\frac{d0}{dt}\right)$$

Collecting results we have

$$v_y = \frac{dy}{dt} \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}45)$$

$$v_r = \frac{dr}{dt} \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}46)$$

$$v_\theta = r\frac{d\theta}{dt} \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}47)$$

$$f_y = m\frac{d^2y}{dt^2} \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}48)$$

$$f_r = m\left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right] \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}49)$$

$$f_\theta = \frac{m}{r}\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) \text{ or } \frac{m}{r}\left(r^2\frac{d^2\theta}{dt^2} + 2r\frac{dr}{dt}\cdot\frac{d\theta}{dt}\right) \quad . \quad (3\text{–}50)$$

In the case under consideration $f_y = 0$ and $f_\theta = Hqdr/dt$. The radial force acting outwardly on the electron is Xq and in addition to this is a radial inwardly directed force due to the magnetic field and the transverse velocity component given by (3–47). Hence the resultant radial force is $Xq - Hq\,rd\theta/dt$, and we may write

$$\frac{m}{r}\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) = Hq\frac{dr}{dt} \quad . \quad . \quad . \quad (3\text{–}51)$$

and

$$m\left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right] = Xq - Hq\,r\frac{d\theta}{dt} \quad . \quad . \quad (3\text{–}52)$$

Integrating (3–51)

$$m\int_{r_f}^{r} d\left(r^2\frac{d\theta}{dt}\right) = Hq\int_{r_f}^{r} rdr$$

$$m\left[r^2\frac{d\theta}{dt} - r_f^2\left(\frac{d\theta}{dt}\right)_f\right] = \tfrac{1}{2}Hq(r^2 - r_f^2)$$

which leads to

$$\frac{d\theta}{dt} = \frac{Hq}{2m}\left(1 - \frac{r_f^2}{r^2}\right) + \frac{r_f^2}{r^2}\left(\frac{d\theta}{dt}\right)_f \quad . \quad . \quad (3\text{–}53)$$

where the suffix f denotes corresponding values of r and $d\theta/dt$. Substituting (3–53) in (3–49) results in

$$m\left\{\frac{d^2r}{dt^2} - r\left[\frac{Hq}{2m}\left(1 - \frac{r_f^2}{r^2}\right) + \frac{r_f^2}{r^2}\left(\frac{d\theta}{dt}\right)_f\right]^2\right\}$$

$$= Xq - Hqr\left[\frac{Hq}{2m}\left(1 - \frac{r_f^2}{r^2}\right) + \frac{r_f^2}{r^2}\left(\frac{d\theta}{dt}\right)_f\right] \quad . \quad (3\text{–}54)$$

Now for practical purposes this expression may be simplified by the assumptions that r_f is negligible compared with r and that the electrons are emitted with negligible velocities. In these circumstances (3–54) reduces to

$$m\left[\frac{d^2r}{dt^2}-\left(\frac{Hq}{2m}\right)^2 r\right]=Xq-\frac{H^2q^2r}{2m}$$

Multiplying throughout by dr/dt and integrating

$$\int_{r_f}^{r}\frac{d^2r}{dt^2}\,dr=\int_{r_f}^{r}\frac{Xq}{m}\,dr-\frac{H^2q^2}{4m^2}\int_{r_f}^{r}r\,dr$$

or

$$\frac{1}{2}\left[\left(\frac{dr}{dt}\right)^2-\left(\frac{dr}{dt}\right)^2\right]=\frac{Vq}{m}-\frac{H^2q^2}{8m^2}(r^2-r_f^2)$$

where $V=\int_{r_f}^{r}X\,dr$

Again, assuming $(dr/dt)_f$ and r_f to be negligible, we have

$$\left(\frac{dr}{dt}\right)^2=\frac{2Vq}{m}-\frac{H^2q^2}{4m^2}r^2$$

A matter of particular interest in connexion with this last equation is that if $V < H^2qr^2/8m$, dr/dt is imaginary. Physically, this means that electrons will never reach the cylinder, but will be subsequently returned to the filament. Hence the smallest value that V may have in order that electrons may arrive at the cylinder is

$$V=\frac{H^2q}{8m}r_c^2$$

Under these conditions $dr/dt = 0$, the electron trajectory just grazes the cylinder and is tangential to this thereat.

The foregoing type of electron motion will again be referred to in dealing with the magnetron in Chapter XII.

Electron Theory of Physical Constants

By means of the concept of the electron it is possible to give an explanation of the principal electrical and magnetic properties of matter such as specific inductive capacity, susceptibility, conductivity, etc. The present chapter will be concluded with an account of the first two mentioned properties, while the third property will be dealt with in the ensuing chapter.

SPECIFIC INDUCTIVE CAPACITY. REFRACTIVE INDEX

Imagine a lamina of non-conducting material in which an electric field has been produced. The electron shells of the atoms are not rigidly attached to their nuclei, so that it is probable that the field produces a relative displacement of the two. As the electronic charge is negative, it follows that the displacement of the shells is in the opposite direction to that of the field. A result of this displacement is to produce a number of dipoles within the material, the moment of a dipole being equal to the magnitude of the charge displaced multiplied by the amount of displacement. As a consequence of the dipoles, opposite electric charges appear on the two opposing surfaces of the lamina, the amount of the charge displaced per unit area being termed the polarization and denoted by the vector P_d. If, initially, it is assumed that the atoms, or molecules, constituting the material are sufficiently far apart to be without electrical influence on each other (this is the case with materials of low density), and, further, that the molecules possess no inherent dipole moment, the field within the material is

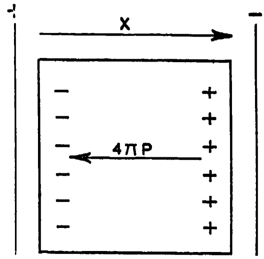


Fig. 3–17

$$X = X_0 - 4\pi P_d \qquad . \quad . \quad . \quad (3\text{–}55)$$

where X_0 is the field in the absence of the material. Referring to Fig. 3–17, which represents the conditions, it is evident that the dielectric polarization sets up a field which opposes that of the original field. Now the specific inductive capacity, or dielectric constant, of a material is defined as the ratio of the field strength due to a given charged system, where there is no material substance (i.e. a vacuum) to the field strength due to the same charge in the same region when occupied by that material. Hence

$$K = \frac{X_0}{X}$$

where K is known as the dielectric constant.

It will now be assumed that when the field displaces an electron it brings into existence a restoring force proportional to the electron displacement. The force on an electron being Xe, we have

$$Xe = \alpha x$$

where x is the displacement and α is a constant. If N_d is the number of electrons displaced per unit volume of the material and has the same value for all of these, then the number per unit area on the surface of the lamina is $N_d x$ and $P_d = N_d ex$. Hence, dividing (3–55) by X,

$$\frac{X_0}{X} = K = 1 + \frac{4\pi P_d}{X} \qquad . \quad . \quad . \quad (3\text{–}56)$$

$$= 1 + \frac{4\pi N_d ex}{\frac{\alpha x}{e}}$$

$$= 1 + \frac{4\pi N_d e^2}{\alpha}$$

In the case of a gas, N_d is proportional to the density, ρ, of the gas. Thus,

$$K - 1 \propto \rho$$

and this relation has been found to hold very accurately.

The refractive index of a medium is given by $n = c/v$ where c is the velocity of light *in vacuo* and v its velocity in the medium. According to Maxwell's electromagnetic theory

$$c = \frac{1}{\sqrt{K_0\mu_0}} \text{ and } v = \frac{1}{\sqrt{K\mu}}$$

and

$$n = \sqrt{\frac{K\mu}{K_0\mu_0}}$$

where K and μ are, respectively, the specific inductive capacity and permeability of the medium and K_0 and μ_0 are those for a vacuum. For most dielectrics μ is practically equal to μ_0 and taking K_0 as 1

$$n = \sqrt{K} \text{ and } n^2 = K$$

Thus

$$n^2 - 1 = 4\pi N_d e^2/\alpha$$

REFRACTION AND DISPERSION

The foregoing treatment of specific inductive capacity and refractive index has been carried out for the case of a steady field. The circumstances will now be considered when an electromagnetic wave falls upon a dielectric material. The electron is now subjected to an alternating electric field which for a given frequency may be written $X \sin pt$, where $p = 2\pi f$, f being the frequency. Assuming the nuclei to remain at rest, the electrons will execute forced harmonic vibrations, the equation of motion being

$$m\frac{d^2x}{dt^2} + \beta\frac{dx}{dt} + \alpha x = eX \sin pt \quad . \quad . \quad (3\text{–}57)$$

where β is a damping coefficient. The general solution of (3–57) is

$$x = \varepsilon^{-\frac{\beta}{2m}t}\left(A \sin \sqrt{\frac{\alpha}{m} - \frac{\beta^2}{4m^2}}\,t + B \cos \sqrt{\frac{\alpha}{m} - \frac{\beta^2}{4m^2}}\,t\right)$$

$$+ \frac{eX}{\sqrt{m^2(p_1{}^2 - p^2)^2 + \beta^2p^2}} \sin (pt - \phi) \text{ where } \tan \phi = \frac{\beta p}{m(p_1{}^2 - p^2)}$$

The first term of this represents a damped oscillation having a frequency approximately equal to the free vibration of the electron, i.e. $p_1 = \sqrt{\alpha/m}$, where $p_1 = 2\pi f_1$, f_1 being the natural frequency. This term ultimately dies away. The second term represents a forced harmonic vibration and hence, after the transient has disappeared, we may write

$$x = \frac{eX}{\sqrt{m^2(p_1{}^2 - p^2)^2 + \beta^2p^2}} \sin (pt - \phi) \quad . \quad (3\text{–}58)$$

the amplitude of vibration of the electron being

$$\frac{eX}{\sqrt{m^2(p_1{}^2 - p^2)^2 + \beta^2p^2}} \quad . \quad . \quad . \quad (3\text{–}59)$$

Now, if p_1 is very large, the amplitude of vibration of the electron due to the incident wave is correspondingly small and, thus, it scarcely affects the propagation of the wave. This also tends to be the case when p is large compared with p_1. If $p = p_1$ a condition of resonance results and the energy of the wave is completely absorbed and scattered in every direction. If p is very large it will be noted that the amplitude of vibration is again small, but in this case β^2p^2 is large and the energy of the wave is now largely employed in heating the dielectric.*

The influence of the electronic motion on the dielectric constant and refractive index may now be considered. Writing X sin pt and X_0 sin pt instead of X and X_0 in (3–55), also (3–58) for x in the term containing P_d, we have

$$X \sin pt = X_0 \sin pt - \frac{4\pi N_d e^2 X}{\sqrt{m^2(p_1{}^2 - p^2)^2 + \beta^2p^2}} \sin (pt - \phi)$$

and, if ϕ is small, this leads to

$$n^2 = K = 1 + \frac{4\pi N_d e^2}{\sqrt{m^2(p_1{}^2 - p^2)^2 + \beta^2p^2}} \quad . \quad (3\text{–}60)$$

---

* This, of course, is the principle of H.F. Dielectric Heating.

This last result indicates why light is dispersed when it enters a refracting medium, for the index of refraction depends on the frequency of the incident wave. When p is small, (3–60) reduces to (3–56), which is the refractive index for low frequencies or long waves. As p increases towards p_1, the refractive index increases, becoming a maximum for $p = p_1$. If p now increases beyond p_1 the refractive index may suddenly change sign.* If p is continually increased, the refractive index will, for some value of p, be positive and less than unity and ultimately tend to unity for very short waves. A value of n^2 less than unity indicates that long waves are refracted more than short ones, and this phenomenon has been observed and termed *anomalous* dispersion.

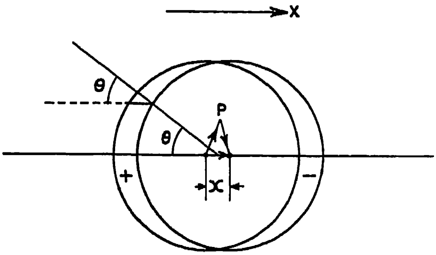


Fig. 3–18

The foregoing theory of specific inductive capacity and refractive index has been based on the assumption that the molecules are sufficiently far apart to be without influence on each other. In the case of dense substances, however, it is necessary to take into consideration the mutual interaction of the dipoles. Previously it was assumed that the field at an atom or molecule is the macroscopic field X. Due to the influence of neighbouring electrons and nuclei, the field at a molecule may differ from X and the actual field is termed the microscopic field. It is the latter which we must now determine. In order to calculate this, the small space surrounding a molecule, in which there are no other charges, may be regarded as a small spherical cavity within the medium with the molecule at its centre. Now the medium may be regarded as consisting of two volume densities of charge $+ \rho$ and $- \rho$ which, in the absence of the macroscopic field X, coincide, giving an electrically neutral condition. When X is applied, the positive electricity is displaced relatively to the negative in the direction of the field through a distance x such that $P_d = \rho x$. The condition of two spheres of positive and negative electricity within the medium, initially coinciding and corresponding to the cavity above, may now be represented by Fig. 3–18, the distance between the centres of the spheres being x. It will be appreciated that surface charges now appear on the walls of the cavity and that the field resulting from these is in the same direction as that of X. If the displacement x is small, the surface density of charge is given by $\rho x \cos\theta$. The field within the cavity resulting from its surface charges is simply that due to the two spheres of positive and negative electricity. At any point P the field due to the former is $\frac{4}{3}\pi r^3\rho/r^2 = 4\pi\rho r/3$, and, similarly, due to the latter, $-4\pi\rho r/3$. The resultant of these will be seen to be $4\pi\rho x/3$, i.e. the field is constant in magnitude and direction throughout the cavity. Hence the microscopic field is

---

*Providing β is zero.

$$X + 4\pi\rho x/3 = X + 4\pi P_d/3$$

and

$$e\left(X + \frac{4\pi P_d}{3}\right) = \alpha x \quad . \quad . \quad . \quad . \quad (3\text{–}61)$$

From (3–56) we have

$$K - 1 = \frac{4\pi N_d e x}{X}$$

and, substituting for x,

$$K - 1 = \frac{4\pi N_d e^2}{\alpha X}\left(X + \frac{4\pi P_d}{3}\right)$$
$$= \frac{4\pi N_d e^2}{3\alpha}\left(3 + \frac{4\pi P_d}{X}\right)$$

But $4\pi P_d/X = K - 1$, so that

$$\frac{K-1}{K+2} = \frac{n^2-1}{n^2+2} = \frac{4\pi N_d e^2}{3\alpha}$$

Hence for denser mediums, the dispersion formula of (3–60) becomes

$$\frac{n^2-1}{n^2+2} = \frac{4\pi N_d e^2}{3\alpha\sqrt{m^2(p_1{}^2 - p^2)^2 + \beta^2 p^2}}$$

PARAELECTRICITY

So far it has been assumed that the polarization of a dielectric medium is due to the relative displacement of the electron shells and the nucleus. In addition to the dipoles produced by this displacement, it is possible for a molecule to possess an inherent dipole moment. Such moments are possessed by molecules only, since (in the absence of an external field) for atoms the centres of gravity of the positive and negative charges coincide. Under the influence of an external field, the permanent dipoles tend to become oriented with their axes in the direction of the field, the orientation being, however, opposed by the thermal agitation of the dielectric. The theory of this paraelectric effect, as it is termed, is exactly similar to that of paramagnetism described on page 144. We may thus apply the results obtained from a study of paramagnetism to electric dipoles and have, therefore, for the paraelectric contribution to the polarization

$$P_p = \frac{N_p p^2}{3kT}\left(X + \frac{4\pi}{3}P\right)$$

where p is the electric dipole moment. The quantity $N_p p^2/3kT$ is known as the paraelectric susceptibility of the dielectric medium. As the dielectric and paraelectric polarizations have the same sign, the total polarization is

$$P = P_d + P_p$$

Taking the mutual interaction of the permanent dipoles into consideration, both these and the dielectric dipoles will contribute to the microscopic field, which will be given by

$$X + \frac{4\pi}{3}(P_d + P_p)$$

Hence we have

$$X = (X_0 - 4\pi P) = X_0 - 4\pi\left\{N_d e x + \frac{N_p p^2}{3kT}\left[X + \frac{4\pi}{3}(P_d + P_p)\right]\right\}$$

Substituting for x from (3–61) and dividing throughout by X, there results

$$1 = K - 4\pi\left\{\frac{N_d e^2}{3\alpha}\left[3 + \frac{4\pi}{X}(P_d + P_p)\right] + \frac{N_p p^2}{9kT}\left[3 + \frac{4\pi}{X}(P_d + P_p)\right]\right\}$$

But $4\pi(P_d + P_p)/X = K - 1$, so that

$$\frac{K-1}{K+2} = \frac{n^2-1}{n^2+2} = \frac{4\pi}{3}\left(\frac{N_d e^2}{\alpha} + \frac{N_p p^2}{3kT}\right)$$

In this expression it will be noted that the paraelectric contribution is temperature dependent. Thus, by measuring K as a function of T, both $N_d e^2/\alpha$ and $N_p p^2/3k$ may be determined. It may be mentioned that the paraelectric contribution is responsible for the high value of the dielectric constant of water.

MAGNETISM

As an electron in motion constitutes a current element and associated with the latter is a magnetic field, it is rational to believe that the magnetic properties of materials are due to electron motions within the atom. These motions may be either due to electrons describing orbits, spinning upon axes through their centres, or combinations of both.

Consider an electron describing an orbit of radius r in time T. We have

$$ids = ev$$

Assuming the orbit to be circular, the distance once round the orbit is $2\pi r$. Hence $v = 2\pi r/T$ and

$$ids = \frac{e2\pi r}{T} = \frac{e}{T} \cdot 2\pi r$$

Thus the motion of the electron in the orbit may be considered to constitute a current of magnitude e/T e.m.u., $2\pi r$ being, of course, equivalent to ds. If the orbit is elliptical the result is similar, for the mean velocity of the electron is still given by ds/T. Now the magnetic moment of a plane current-carrying circuit is equal to the product of the circuit area and the current. Thus the magnetic moment of an electron describing an orbit of area a is ea/T and is perpendicular to the plane of the orbit.

DIAMAGNETISM

Substances which, when placed in a magnetic field, produce a moment opposite to that field are said to be diamagnetic. Obviously the permeability of such substances is less than unity. Diamagnetic substances are composed of atoms whose resultant magnetic moment is zero; i.e. such atoms possess several electrons and orbits and the magnetic moments of these mutually compensate each other, resulting in the magnetic moment of each atom being zero. The cause of diamagnetism arises from the fact that when an atom with its system of orbits is placed within a field H, the whole system undergoes a procession about the field axis. Following Lenz's Law, this procession must create a current whose magnetic moment is such as to oppose H, thus decreasing μ below unity.

To consider this phenomenon in detail, let an electron rotate in a circular orbit of radius r with angular velocity w, the plane of the orbit being resident in a uniform magnetic field H, as shown by Fig. 3–19. Here $ABCD$ is the orbit and DC the field axis. As the moment of the electron in the orbit is ea/T, the torque produced by the field on the orbit is Hea/T or $eawH/2\pi$, as $1/T = w/2\pi$. Now this torque is equal to the reactionary torque of precession, the value of the latter being Iww', where I is the moment of inertia of the precessing body (i.e. the electron) and w' the angular velocity of precession.* Hence

$$\frac{eawH}{2\pi} = Iww'$$

$$= \frac{e\pi r^2 wH}{2\pi} = mr^2ww'$$

and

$$w' = \frac{eH}{2m} \qquad \qquad . \quad . \quad . \quad . \quad . \quad (3\text{–}62)$$

as $a = \pi r^2$ and $I = mr^2$. Thus, the electron orbit precesses about the axis DC with an angular velocity given by $w' = eH/2m$.

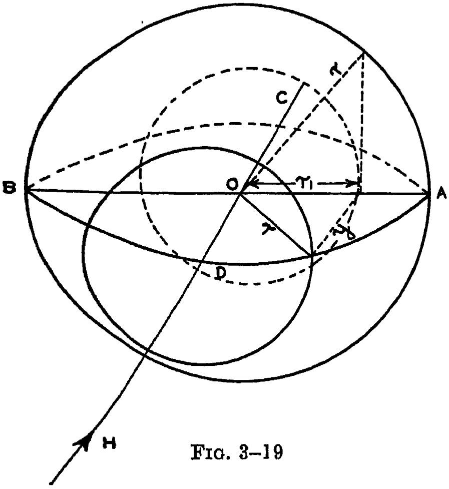


FIG. 3–19

Because of the precessional motion the electron will, when viewed along DC, appear to pursue an orbit of periodically varying area. For example, if the electron were stationary at A or B the area of the orbit would be equal to πr^2. If the electron remained at D the area would be zero. Thus, the magnetic moment of the orbit whose axis is parallel to H is continually varying and it is necessary to determine the average value of this.

Referring to Fig. 3–19, if r_1 is the radius of the precessional orbit as viewed along DC, then the mean value of $r_1{}^2$ is

$$\bar{r}_1{}^2 = \frac{1}{r}\int_0^r r_1{}^2 dy$$

$$= \frac{1}{r}\int_0^r (r^2 - y^2)dy$$

$$= \tfrac{2}{3}\, r^2$$

Thus the mean area of the precessional orbit is

$$\tfrac{2}{3}\, \pi r^2 = \tfrac{2}{3}\, a$$

---

* *The General Properties of Matter*, Newman and Searle, p. 79.

and the moment due to this is

$$M_{dH} = -\tfrac{2}{3}\, ea \,.\, \frac{w'}{2\pi}$$

the negative sign indicating that M_d opposes H. But, from (3-62), $w' = eH/2m$, and hence

$$M_{dH} = -\frac{e^2r^2}{6m} H$$

The average diamagnetic moment for the orbit per unit field strength is then

$$M_d = -\frac{e^2}{6m} r^2$$

This result is, of course, for a single orbit only. If an atom has a number of orbits, then

$$M_{dl} = -\frac{e^2}{6m} \Sigma r_l^2$$

Where Σr_l^2 represents the sum of the squares of the radii of the l orbits.

Now $$M_{dH} = -\frac{I}{N} = -\frac{KH}{N}$$

where I is the intensity of magnetization, i.e. the magnetic moment per unit volume, N the number of atoms per unit volume, and K is the susceptibility per unit volume. Thus

$$-\frac{K}{N} = -\chi = -\frac{e^2}{6m} \Sigma r_l^2$$

where χ is termed the *atomic susceptibility.*

It must be mentioned that diamagnetism is present in all atoms. However, when paramagnetic moments are present, the diamagnetic moment is almost completely masked. This may be demonstrated at once by comparing the diamagnetic and paramagnetic moments for a single orbit. As we have already seen, the paramagnetic moment is ea/T or

$$M_p = \frac{ew\pi r^2}{2\pi}$$

and $$M_d = \frac{e^2r^2}{6m} \text{ (ignoring the negative sign)}$$

Hence $$M_d/M_p = e/3mw$$

and, from Bohr's theory, putting $w = nh/2\pi mr^2$,

$$\frac{M_d}{M_p} = \frac{2\pi}{3}\frac{er^2}{nh}$$
$$= \frac{1{\cdot}4 \times 10^{-10}}{n}$$

for unit field strength. As the diamagnetic moment is proportional to H, the last expression must be multiplied by H if H differs from one. However, even if $H = 10{,}000$, with $n = 1$, M_d/M_p is still only $1{\cdot}4 \times 10^{-6}$.

PARAMAGNETISM

Paramagnetism may be explained on the assumption that atoms have permanent magnetic moments. The magnetic moment of an atom is the vector sum of the magnetic moments due to the orbital and spin motions of its electrons. In the absence of an external field, a substance is paramagnetic when this vector is different from zero, and diamagnetic when equal to zero. According to Langevin's theory of a paramagnetic gas, each gas molecule is supposed to possess a constant magnetic moment M and the distance between molecules to be such that their mutual influence may be neglected. The presence of an external field tends to orient the magnetic moments in the direction of the field, this orientation, however, being opposed by the thermal agitation of the gas. If θ is the angle which the magnetic axis of a molecule makes with the direction of the external field, then $\Sigma M \cos\theta$ is the resultant magnetic moment per unit volume, the sign Σ indicating the summation of the values of $\cos\theta$ for all the molecules in unit volume. It is desired to find $\Sigma M \cos\theta$ as a function of H and T, where the latter are, respectively, the external field strength and absolute temperature.

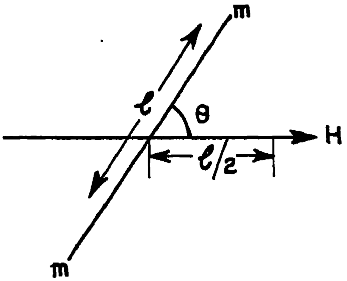


FIG. 3–20

This problem is approached by first considering the *potential* energy of a gas in a gravitational field. If y is the vertical ordinate, then

$$dp = -\rho g dy$$

where p and ρ are, respectively, the gas pressure and density at y.

The negative sign denotes a fall in pressure with increase in y.

We have $$\frac{dp}{dy} = -\rho g \qquad . \quad . \quad . \quad . \quad (3\text{–}63)$$

From Boyle's law $$p\rho_0 = \rho p_0 \qquad . \quad . \quad . \quad . \quad (3\text{–}64)$$

where ρ_0 and p_0 are the values of ρ and p when $y = 0$. From (3–63) and (3–64)

$$\frac{dp}{p} = \frac{\rho_0}{p_0} g dy$$

and $$p = C\varepsilon^{-\frac{\rho_0}{p_0} gy}$$

where C is a constant.

When $y = 0$, $p = p_0$, from which

$$p = p_0\varepsilon^{-\frac{\rho_0}{p_0} gy}$$

From the kinetic theory of gases

$$\frac{p}{p_0} = \frac{n}{n_0}$$

$$\frac{\rho_0}{p_0} = \frac{m}{kT}$$

Hence $$\frac{n}{n_0} = \varepsilon^{-\frac{gm}{kT}}$$

where n is the number of molecules per c.c., m the mass of a molecule, T the absolute temperature, and k is Boltzmann's constant. Now gmy is the potential energy of a molecule at y and, writing this as w,

$$\frac{n}{n_0} = \varepsilon^{-\frac{w}{kT}}$$

Thus, as n_0 is constant, it follows that the number of molecules having potential energy w in a medium consisting of molecules having kinetic energy $\frac{3}{2}$ kT, is proportional to $\varepsilon^{-\frac{w}{kT}}$

Referring to Fig. 3–20, it will be seen that the potential energy of an elementary magnet of moment ml, the axis making an angle θ with H, is

$$2mH\left(\frac{l}{2} - \frac{l}{2}\cos\theta\right)$$
$$= MH(1 - \cos\theta)$$

this expression giving the work done in turning the magnetic axis from a position parallel to the field to that shown. It follows by an analogy with the gravitational potential of molecules in a gravitational field, that the number of molecules having magnetic potential energy $MH(1 - \cos\theta)$ in a magnetic field is proportional to

$$\varepsilon^{\frac{-MH(1-\cos\theta)}{kT}}$$

Hence
$$dn = C\varepsilon^{\frac{MH\cos\theta}{kT}} \sin\theta d\theta \quad . \quad . \quad . \quad . \quad (3\text{–}65)$$

where C is independent of θ and contains $\frac{MH}{kT}\varepsilon^{-\frac{MH}{kT}}$

Putting $\cos\theta = x$, $dx = -\sin\theta d\theta$.

Substituting and integrating

$$n = C\int_{-1}^{+1} \varepsilon^{\frac{MHx}{kT}} dx$$

from which
$$C = \frac{-nMH}{kT(\varepsilon^{MH/kT} - \varepsilon^{-MH/kT})}$$

It follows that

$$\Sigma M\cos\theta = I = -\int_0^{\pi} \frac{nM^2H\varepsilon^{MH\cos\theta/kT}\cos\theta\sin\theta d\theta}{kT(\varepsilon^{MH/kT} - \varepsilon^{-MH/kT})}$$

where I is the intensity of magnetization, or magnetic moment per unit volume. Integrating

$$I = nM\left[\frac{(\varepsilon^{MH/kT} + \varepsilon^{-MH/kT})}{(\varepsilon^{MH/kT} - \varepsilon^{-MH/kT})} - \frac{kT}{MH}\right]$$
$$= nM\left(\coth\frac{MH}{kT} - \frac{kT}{MH}\right)$$

when MH/kT is large, at very low temperatures or high values of H, for example, $I = nM = I_0$

This corresponds to a condition of saturation when all the atomic magnetic moments have the same direction as the field. Again, considering the case when MH/kT is small, we may write

$$\frac{I}{nM} = \frac{\varepsilon^a + \varepsilon^{-a}}{\varepsilon^a - \varepsilon^{-a}} - \frac{1}{a} \quad . \quad . \quad . \quad (3\text{–}66)$$

where $a = MH/kT$. Expanding ε^a as a power series and neglecting powers higher than a, gives

$$\frac{I}{nM} = \frac{a}{3} = \frac{MH}{3kT} \quad . \quad . \quad . \quad . \quad (3\text{–}67)$$

from which
$$I = \frac{nM^2H}{3kT}$$

and
$$\kappa = \frac{nM^2}{3kT}$$
where κ is the magnetic susceptibility per unit volume. This result is in accordance with experiment.

In developing the foregoing theory of paramagnetism the mutual influence of the magnetic dipoles has been ignored as initially was that of the electric dipoles in dealing with refractive index. In the magnetic case, I corresponds to P in the electric case. Thus, the interaction of the magnetic dipoles may be taken into account by putting the microscopic field equal to the macroscopic field H, plus a second quantity proportional to the magnetization I. Experiment shows, however, that $4\pi/3$ gives incorrect results when taken as the coefficient of I and so a general empirical constant b is taken instead. Hence
$$I = \frac{nM^2}{3kT}(H + bI)$$
Solving for I we obtain
$$I = \frac{nM^2H}{3k\left(T - \dfrac{bnM^2}{3k}\right)} \quad . \quad . \quad . \quad (3\text{–}68)$$
or
$$\kappa = \frac{K}{3k(T - \phi)} \quad . \quad . \quad . \quad . \quad (3\text{–}69)$$
The temperature ϕ is known as the Curie point.

The relation expressed by (3–69) shows that if ϕ is positive there should be some temperature at which the susceptibility attains an extremely large value (theoretically, infinite). This is the case for a small group of substances, namely the ferromagnetic metals. However, in this case a may not be small, and hence the appropriate formula is (3–66) rather than (3–69). The value of a, however, in (3–66) must be taken as mH'/kT where $H' = (H + bI)$. A further important point concerning ϕ is that below this temperature a substance may be magnetized even if $H = 0$. To determine I in these circumstances we have the three equations
$$\frac{I}{I_0} = \left(\coth a - \frac{1}{a}\right) \quad . \quad . \quad . \quad (3\text{–}70)$$
$$I = \frac{H' - H}{b} \quad . \quad . \quad . \quad . \quad (3\text{–}71)$$
$$a = \frac{MH'}{kT} \quad . \quad . \quad . \quad . \quad . \quad (3\text{–}72)$$

From (3–71) and (3–72)

$$I = \frac{akT/M - H}{b}$$

and if

$$H = 0$$

$$I = \frac{akT}{bM}$$

or

$$\frac{I}{I_0} = \frac{akT}{bnM^2}$$

But from (3–68)

$$b = \frac{3k\phi}{nM^2}$$

and hence

$$\frac{I}{I_0} = \frac{aT}{3\phi} \qquad . \quad . \quad . \quad . \quad . \quad (3\text{–}73)$$

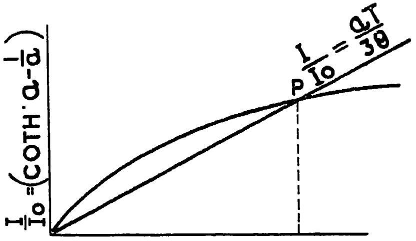


FIG. 3–21

Plotting (3–70) and (3–73), Fig. 3–21, the latter intersects the former at O and P. Hence I/I_0 may have either the value zero or PP_1. In the event of the latter value obtaining, the substance is spontaneously magnetized to an amount given by P from which it is evident that what is known as remanence (or residual magnetism) is due to the molecular field originating from the magnetic dipoles. There is no spontaneous magnetization unless the line OP lies below the curve, and this only occurs when the slope of the line is less than that of (3–70) at the origin. From (3–67) it will be seen that the slope must be less than 1/3, i.e.

$$\frac{T}{3\phi} < \frac{1}{3}$$

or

$$T < \phi$$

which result has already been indicated by the approximate formula of (3–69). Above the temperature ϕ, ferromagnetic substances become paramagnetic, and hence ϕ denotes a transition temperature.

CHAPTER IV

ELECTRONS IN METALS

Classical Theory of Conductivity

THAT outstanding class of substances known as metals is particularly distinguished by reason of high thermal and electrical conductivity and reflexion of radiation. The explanation of such properties was difficult, if not impossible, prior to the discovery of the electron, but with the advent of the latter not only could these properties be explained, but many other phenomena associated with the metallic state. The initial clarification of this important subject took place principally at the hands of Drude and Lorentz, who introduced what may be termed the electron theory of metals. This theory followed classical lines in that the electrons, etc., within a metal were considered to act in accordance with Newtonian Laws. Subsequent investigation has shown that the state of electrons within a metal is not in accordance with classical theory, yet, nevertheless, the results which may be derived from classical assumptions yield, in many cases, information which is closely in accordance with fact. Hence, initially, the electronic state of a metal will be considered according to theories initiated by Lorentz.

The relatively high electrical and thermal conductivities of metals may be explained on the supposition that within a metal is a number of free electrons. These may be supposed to possess a random motion, similar to the molecules of a gas and, as as many move in one direction as in another, the average current, in the absence of a field, is zero. On applying a potential difference between the ends of a conductor an electric field is produced and the electrons are set in motion in the direction of the field. The atoms (or ions) constituting the metal are bound within the lattice structure, and hence, except perhaps for a small initial displacement, remain at rest.

Assuming the electrons to behave as a gas, each, on the average, will travel a distance l, the mean free path, between collisions. It may also be assumed that, as after a collision with an atom the probability of an electron having any direction is the same, the average velocity of a large number of electrons which have just experienced a collision is zero. Under the influence of the electric field, X, each electron will experience a force eX and will be accelerated by this force for a period equal to that necessary to describe the mean free path l. This time is given by $l/\bar{C}$, where $\bar{C}$ is the mean random velocity of the electrons. Hence the electron velocity added by the field at the end of this time is

$$\frac{eXl}{m\bar{C}}$$

and the mean velocity is

$$v = \frac{eXl}{2m\bar{C}}$$

The charge conveyed by the electrons per unit cross-section of the conductor per second is

$$nev = \frac{ne^2Xl}{2m\bar{C}}$$

where n is the number of electrons per unit volume. But this is the current density J. Thus

$$J = \frac{ne^2Xl}{2m\bar{C}} = \sigma X$$

As none of the factors composing σ depends on X, $J \propto X$ and we have an explanation of Ohm's Law.

The expression for the current density may be put in a different form. According to the law of equipartition of energy

$$\tfrac{1}{2}m\left(\sqrt{\frac{3\pi}{8}}\,\bar{C}\right)^2 = \tfrac{3}{2}\,kT$$

and thus

$$\sigma = \frac{\pi n e^2 l\bar{C}}{16kT} \quad . \quad . \quad . \quad (4\text{–}1)$$

Now n is likely to be independent of temperature, while it is known from experiment that the conductivity of all pure metals is inversely proportional to the absolute temperature. From this it may be concluded that $l\bar{C}$ is independent of temperature. Qualitatively it may be seen that temperature affects each of these two quantities oppositely. An increase in temperature causes the lattice elements to vibrate with greater amplitude, this increasing the collision frequency between atoms and electrons and causing a reduction in l.

THERMAL CONDUCTIVITY

Consider a plane AB in a conductor and two others, C, D, parallel to it and at distances from AB equal to the mean free path of the electrons, Fig. 4–1. The number of electrons passing through unit area of each side of AB per second, is, according to the kinetic theory,

$$n\sqrt{\frac{kT}{2\pi m}}$$

If the temperature is everywhere the same, these two groups carry the same energy and thus the net energy transfer across AB is zero. Now let the temperature at C be T_1 and at D, T_2, and $T_1 > T_2$. In these circumstances there will be an energy transfer across AB from C to D. The energy transferred per second from C to D is

$$n\sqrt{\frac{kT}{2\pi m}} \cdot \frac{3}{2}kT_1$$

and from D to C $\quad n\sqrt{\frac{kT}{2\pi m}} \cdot \frac{3}{2}kT_2$

Therefore the net energy transfer per second across AB from C to D is

$$\frac{3}{2}nk\sqrt{\frac{kT}{2\pi m}}(T_1 - T_2)$$

where T is the temperature at AB.

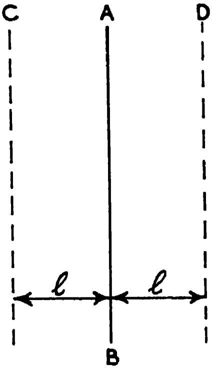


FIG. 4–1

The temperature gradient at AB is $(T_1 - T_2)/2l$ and, if L is the coefficient of thermal conductivity, the quantity of heat transferred across unit area of AB per second is

$$L = \frac{T_1 - T_2}{2l} = \frac{3}{2}nk\sqrt{\frac{kT}{2\pi m}}(T_1 - T_2)$$

or $$L = 3nkl\sqrt{\frac{kT}{2\pi m}}$$

Now $$\tfrac{1}{2}m\left(\sqrt{\frac{3\pi}{8}}\,\bar{C}\right)^2 = \tfrac{3}{2}kT$$

and $$L = \tfrac{3}{4}nkl\bar{C} \quad . \quad . \quad . \quad (4\text{–}2)$$

Again, we do not know the manner in which the separate factors of (4–2) depend on temperature, but dividing (4–2) by (4–1) we have

$$\frac{12}{\pi}\left(\frac{k}{e}\right)^2 T$$

i.e. the ratio of thermal to electrical conductivity is proportional to the absolute temperature, which is the law of Weidemann-Franz.

This law is well substantiated by experiment, the factor of proportionality agreeing with $12/\pi \, . \, (k/e)^2$.

Electromagnetic Induction

The phenomenon of electromagnetic induction may be readily explained by the supposition of free electrons within a metal. For example, let a wire of length l move with velocity v in a direction normal to a field of strength H. The positive ions within the lattice structure of the wire being in motion constitute a current element equal to ev and consequently experience a force, at right angles both to v and H, of magnitude Hev. Because of the restraint imposed by the lattice, the positive ions undergo a small initial displacement only. The electrons experience the same force, but in the opposite direction. Being free to move through the lattice structure, the electrons tend to accumulate at one end of the wire until the electric field, X, which they create produces a force on them which just balances that due to H. The potential difference between the ends of the wire is $V = Xl$ and thus $X = V/l$. The force on an electron is $Xe = Ve/l$ and for equilibrium we have

$$Hev = Ve/l$$

or

$$V = Hvl$$

But vl is rate of change of area A normal to H, and thus

$$V = H\frac{dA}{dt}$$

Also $HdA = d\phi$, where $d\phi$ is the change of flux in time dt, therefore

$$V = \frac{d\phi}{dt}$$

i.e. the rate of cutting lines of force.

Thermionics

If the temperature of an electrically charged body is raised sufficiently, it is found that the charge is lost more rapidly as the temperature increases. This phenomenon is particularly noticeable when the initial charge is a negative one. To such phenomena Richardson applied the term *thermionics*. For the study of thermionics the simple apparatus of Fig. 4–2 may be employed. It consists of a loop of tungsten or platinum wire suspended within a metal cylinder, the whole being enclosed in a highly-evacuated glass bulb. The loop is heated by the passage of an electric current and its temperature found from its resistance or by means of an optical pyrometer. The cylinder is connected to one side of a source of d.c. potential, via a galvanometer or microammeter, while the other side of the supply is connected to one end of the wire loop. On raising the temperature of the loop it is found that a current is registered by the galvanometer when the loop is negative, but not when positive. Hence the apparatus functions as a valve, permitting a flow of negative electricity only from loop to cylinder, when the latter is at a positive potential with regard to the former.

The magnitude of the current obtainable in the foregoing experiment depends on the temperature and the potential difference, tending to increase with an increase in either of these. If the temperature is maintained constant while the potential of the cylinder is continually raised, the current at first rapidly increases but ultimately attains a constant maximum value. This value is known as the saturation current and increases with an increase in temperature.

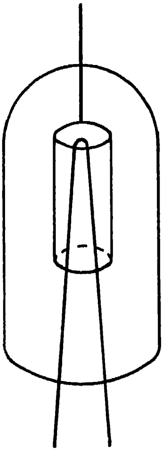


FIG. 4–2

The negative electricity emanating from a hot metal consists of electrons. This was first established by J. J. Thomson, who measured the ratio e/m and found it to be that for the electron. According to the classical theory of thermionic emission, it is considered that the electrons within the heated metal share in its thermal agitation and that the more energetic electrons are evaporated in a similar manner to molecules from the surface of a liquid. A matter of obvious importance is the relation between the thermionic current and emitter temperature. A number of methods of deriving this have been found and we shall first consider those based on the classical theory and afterwards those based on thermodynamics and the quantum theory.

Suppose a large metal plane surface to be raised to such a temperature that electrons escape from it. The space above the surface will be filled with electrons which will behave as a monatomic gas of small molecular weight. The electronic gas pressure is given by $P = kTn_x$, where k is Boltzmann's constant and n_x the number of electrons per unit volume at distance x above the plane. Equilibrium will be attained when the number n_x of electrons per unit volume has reached such a value that the number of electrons which strike the plane and are absorbed by it is equal to the number emitted by it in the same time. The number of x-directed electrons per unit area per second is*

$$n_0 = n_x \sqrt{\frac{kT}{2\pi m}} \quad . \quad . \quad . \quad . \quad (4\text{–}3)$$

this being the number of electrons emitted by unit area of the plane per second.

Now to separate an electron from the surface of the metal requires work to be done. Let X be the force (the nature of this force will be discussed later) on an electron at a distance x from the surface of the metal. This force is zero inside the metal and becomes zero at some distance above it. Due to the force tending to draw the electrons towards the surface, there will be an outward pressure gradient in the electron gas, the force per unit volume of electrons being equal to the pressure gradient, i.e. dP/dx. The attraction of the plane will produce a force on unit volume of the electrons outside the metal equal to Xn. For equilibrium these forces must balance and we have

$$\frac{dP}{dx} = Xn$$

or

$$kT\frac{dn}{n} = Xdx$$

Hence

$$kT[\log_\varepsilon n]_0^{x_1} = \int_0^{x_1} Xdx$$

and

$$kT \log \frac{n_{x_1}}{n} = -\phi e$$

$$n_{x_1} = n\varepsilon^{-\frac{\phi e}{kT}}$$

where x_1 is a distance from the plane at which the force on an electron is zero and ϕe is the work done in removing an electron from the metal surface.

Substituting for n_{x1} from (4–3)

$$n_0 = n\sqrt{\frac{kT}{2\pi m}}\,\varepsilon^{-\frac{\phi e}{kT}}$$

$$= AT^{\frac{1}{2}}\varepsilon^{-\frac{\phi e}{kT}}$$

Since each electron carries a charge e, the thermionic current density is n_0e. The quantity ϕe is termed the thermionic work function and is measured in electron-volts. Consequently, ϕ is expressed in volts.

---

* See p. 172.

A thermodynamical method of deriving the thermionic current/temperature relationship is as follows. Consider an enclosure fitted with a piston, and let the base of the enclosure be the metal emitting electrons. The result of taking the electron atmosphere round a Carnot cycle will be considered. The cycle is started with an infinitesimal adiabatic expansion, the absolute temperature and pressure falling from T and P to $T-dT$ and $P-dP$. This is followed by an isothermal compression, the volume decreasing by dS, during which operation ndS electrons are driven into the metal. An adiabatic compression follows, and then an isothermal expansion until the pressure is again P. As the final temperature and pressure are the same as at the commencement of the cycle, n is also the same. Further, if the cycle is very small it will have the form of a parallelogram. During the isothermal expansion at T the work done is $ndSE + PdS$, where E is the work done in extracting an electron from the metal; i.e. E is the thermionic work function. Now, from the second law of thermodynamics,

$$\frac{\text{work for cycle}}{\text{work absorbed at } T} = \frac{T-(T-dT)}{T}$$

or

$$\frac{dSdP}{ndSE + PdS} = \frac{dT}{T}$$

and

$$TdP = nEdT + PdT$$

The electron gas pressure is

$$P = knT \quad . \quad . \quad . \quad . \quad (4\text{–}4)$$

Differentiating (4–4)

$$dP = kndT + kTdn$$

and thus

$$knTdT + kT^2dn = nEdT + PdT$$

from which

$$\frac{dn}{n} = \frac{E}{kT^2}dT$$

Integrating

$$\log_e n = \int \frac{E}{kT^2}\,dT + c$$

and

$$n = A\varepsilon^{\int \frac{E}{kT^2}dT} \quad . \quad . \quad . \quad . \quad (4\text{–}5)$$

The number of electrons striking unit area of the base of the enclosure per second is

$$n\sqrt{\frac{kT}{2\pi m}}$$

and hence
$$n_0 = A\sqrt{\frac{kT}{2\pi m}}\,\varepsilon^{\int\frac{E}{kT^2}dT}$$
$$= A_1 T^{\frac{1}{2}}\,\varepsilon^{\int\frac{E}{kT^2}dT}$$
which, if E is constant, becomes
$$n_0 = A_1 T^{\frac{1}{2}}\,\varepsilon^{-\frac{E}{kT}} = A_1 T^{\frac{1}{2}}\,\varepsilon^{-\frac{\phi e}{kT}} \quad . \quad . \quad (4\text{–}6)^*$$
the thermionic current density being $n_0 e$.

It will be noted that by the assumption that E is constant, the same result is obtained as by the first method. If the kinetic energy of an electron within the metal be taken as negligible, then E is the work done in extracting it, plus the work done in giving it kinetic energy $\frac{1}{2}mC^2$ in the space outside the metal. Now
$$\frac{1}{2}\,mC^2 = \frac{3}{2}\,kT$$
and hence
$$E = E_0 + \frac{3}{2}\,kT$$
where E_0 is solely the work done in extracting the electron from the metal. In this case
$$\int\frac{E}{kT^2}\,dT = \int\frac{E_0}{kT^2}\,dT + \frac{3}{2}\int\frac{dT}{T}$$
$$= -\frac{E_0}{kT} + \frac{3}{2}\log_\varepsilon T$$
and
$$\varepsilon^{\int\frac{E}{kT^2}dT} = \varepsilon^{-\frac{E_0}{kT}}\,T^{\frac{3}{2}}$$
which leads to
$$n_0 = A_2 T^2 \varepsilon^{-\frac{E_0}{kT}} = A_2 T^2 \varepsilon^{-\frac{\phi_0 e}{kT}} \quad . \quad . \quad (4\text{–}7)$$

A further way of regarding thermionic emission which, however, is similar to the previous thermodynamical treatment, is as follows. Regarding the emitter and electrons as analogous to a liquid and its vapour when in equilibrium, we may apply the latent heat equation, viz.
$$L = T\frac{dP}{dT}(V_1 - V_2) \quad . \quad . \quad . \quad (4\text{–}8)$$
where L is the energy absorbed when one mol of electrons evaporates, V_1 and V_2 are the volumes of one mol of electrons in the space above and in the metal itself, P the electronic gas pressure, and T the absolute temperature. As V_2 is small compared with V_1 (4–8) reduces to
$$L = TV_1\frac{dP}{dT}$$

---

* (4–6) is sometimes referred to as Richardson's Equation.

L is absorbed in two ways, (a) in effecting the escape of electrons through the emitter surface, (b) in overcoming the electronic gas pressure into which the electrons evaporate. The former energy is equal to NE and the latter to PV, where N is Avogadro's number. As $PV = kNT$ we have

$$L = NE + kNT$$

and, from (4-8),

$$TV\frac{dP}{dT} = N(E + kT)$$

or, writing $V = kNT/P$,

$$kN\frac{T^2}{P}\frac{dP}{dT} = N(E + kT)$$

$$k\frac{dP}{P} = \frac{E + kT}{T^2}\,dT$$

$$P = KT\varepsilon^{\int \frac{E}{kT^2}dT}$$

and, as $P = knT$,

$$n = c\varepsilon^{\int \frac{E}{kT^2}dT} \quad . \quad . \quad . \quad (4\text{–}9)$$

where K and c are constants. As (4–9) is identical with (4–5) it obviously leads to the same result as given by (4–6) and (4–7).

The Fermi-Dirac Theory of the Electron Gas

The classical viewpoint of a conducting solid is that of an equipotential volume. In the present book this viewpoint has, so far, been adopted, but it is now necessary to inquire more closely into the conditions existing within conducting solids, such as metals, particularly as these constitute one of the most important sources of electrons. As shown in Chapter V, metals are crystalline in structure and consist of a lattice work of ions which is founded on the constant repetition of a unit cell in three dimensions. Because of the relatively small distance separating atoms in metals, the outer shell, or valance, electrons are no more closely associated with one atom than another, with the result that the attachment of such electrons to any particular atom is practically negligible. The result of this is that at least one, and sometimes two or three, electrons per atom are free to move throughout the interior of the metal, particularly when under the influence of an applied electric field. These electrons are termed *free electrons* and are responsible for the electrical and thermal conductivities of a solid as previously explained.

Due to the lattice structure of ions within a metal it cannot be considered as an equipotential volume, but rather as one in which exist intense local variations in potential. At any given point the potential is that due to the sum of the potentials produced at this point by all the ions. Considering the effect due to one ion, the atomic number of which is Z, the positive charge on the nucleus is Ze and this is surrounded by an approximately spherical cloud of electrons. The potential at a distance r from the nucleus, when r is within the electron cloud, is that due to the charge on the nucleus and is Ze/r. Hence the potential energy of an electron situated at r is $-Ze^2/r$, the negative sign being due to the electron having a negative charge. When r lies outside the electron cloud the potential

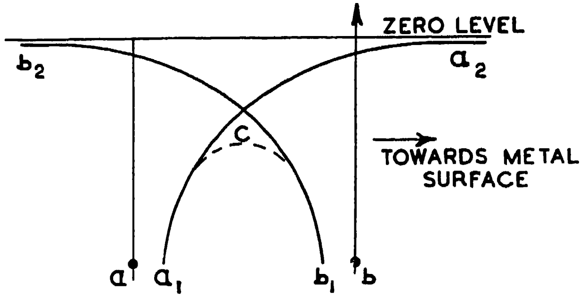


FIG. 4–3

due to an atom from which a planetary electron has become detached is $[Ze - (Z - 1)e]/r = e/r$. The corresponding potential energy is, of course, $-e^2/r$. Thus the potential energy of an electron close to the nucleus is given by $-Ze^2/r$ and remote from the nucleus by $-e^2/r$.

Considering now the potential energy of an electron due to two adjacent ions in the metal lattice, the conditions may be represented by Fig. 4–3. Here a and b represent the nuclei of the two ions, and a_1a_2 and b_1b_2 their respective potential energy curves. Ignoring the effect of other ions, the resultant potential energy curve is given by the sum of the ordinates of these two curves as shown by a_1cb_1. The distance from the nuclei is measured horizontally and the potential energy vertically. Hence the curves relating to the latter are plotted below the abscissa as the potential energy of the free electrons within the metal is negative. It will be noted that in the vicinity of the nuclei the resultant potential energy curve is practically coincident with the component curves, but is lower and possesses a smaller radius of curvature between the nuclei. Considering a row of nuclei, shown in Fig. 4–4, the potential energy distribution may be represented as indicated, the tops of the curves being somewhat flatter than those of Fig. 4–3 due to the influence of other ions. The diagram indicates that the potential tends to be constant over large regions because the volume of matter consists principally of space, that occupied by the nuclei being extremely small.

Passing to a consideration of conditions existing at the surface of a metal, indicated by the line *SS*, it will be noted that no ions exist to the right of this line. Hence, proceeding towards the surface from the last ion of the lattice-structure, there is no depression of the potential-energy curve, as between ions within the metal, but a continual rise in the curve until the surface is reached. Thus it will be appreciated that what may be termed a *potential barrier* exists at the surface of a metal and that this is independent of any potential that we may choose to give to the metal as a whole.

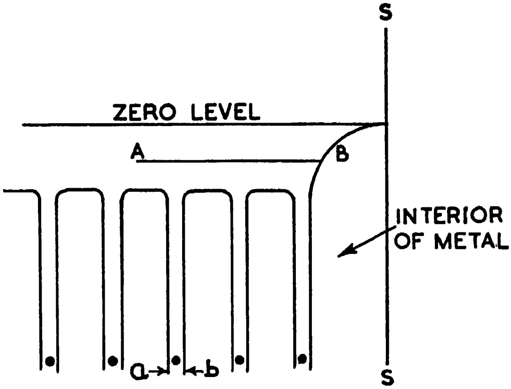


Fig. 4–4

Considering electrons in the vicinity of an ion, these will rebound from the potential walls *ab* and thus be restricted in motion. Such electrons do not contribute to the conductivity of a metal and are termed *bound* electrons. Electrons remote from an ion, such as at *BA*, are free electrons capable of moving freely through the metal. An electron in this category, moving close to the surface, will collide with the potential barrier at *B*, at which point its kinetic energy is entirely converted into potential energy. Hence it will rebound from the potential barrier and no escape from the metal is possible. On the other hand, an electron the kinetic energy of which is greater than that of the potential barrier will make no collision with the latter and may thus escape from the parent metal.

THE SPECIFIC HEAT DILEMMA

It was originally considered that the free electrons within a metal behaved as a monatomic gas obeying gas laws. This view appeared to be supported by the fact that electrons emitted by a hot body possess a Maxwellian distribution of velocities. Now, although the concept of an electron gas accounts for certain phenomena of metallic conduction, some of its consequences nevertheless lead to contradiction of experience. Principal among these is the effect of the contribution of the electrons to the energy content, and hence the specific heat of a metal. According to the kinetic theory of gases, the average kinetic energy of a molecule is $\frac{1}{2}kT$ for each degree of freedom. Hence the energy per gram-mol. is $\frac{1}{2}nRT$ where n is the number of degrees of freedom. For a monatomic gas, $n = 3$, and thus the specific heat per gram-mol. is, in this case, $\frac{1}{2}3RT/T = \frac{3}{2}R$. In the case of a solid (which in a gaseous state is monatomic) the total energy of the molecules is half kinetic and half potential, as will now be shown.

Consider an oscillator of mass m undergoing a simple harmonic motion and subject to no losses. Its energy may be written

$$\tfrac{1}{2}m\dot{x}^2 + \tfrac{1}{2}\mu x^2 = E = \text{constant} \qquad (4\text{–}10)$$

where μ is the force for unit displacement. The mean energy over a cycle is

$$\int_0^T \tfrac{1}{2}m\dot{x}^2 dt + \int_0^T \tfrac{1}{2}\mu x^2 dt = \int_0^T E dt . \qquad (4\text{–}11)$$

where T is the periodic time. Putting

$$\begin{aligned} x &= A \sin pt \\ x^2 &= A^2 \sin^2 pt \\ \dot{x} &= Ap \cos pt \\ \dot{x}^2 &= A^2p^2 \cos^2 pt \end{aligned}$$

where A is the amplitude of the oscillation and $p = 2\pi f$. Substituting in (4–11)

$$\tfrac{1}{2}mA^2p^2\int_0^{2\pi/p} \cos^2 pt dt + \tfrac{1}{2}\mu A^2\int_0^{2\pi/p} \sin^2 pt dt = E$$

and $$\tfrac{1}{4}mA^2p^2 + \tfrac{1}{4}\mu A^2 = E$$

But from (4–10), when $t = 0$

$$\tfrac{1}{2}mA^2p^2 = E$$

Also when $pt = \pi/2$ $\quad \frac{1}{2}\mu A^2 = E$

Hence $\quad \frac{1}{2}mA^2p^2 = \frac{1}{2}\mu A^2$

and $\quad \frac{1}{4}mA^2p^2 = \frac{1}{4}\mu A^2 = \frac{E}{2}$

or the mean value of the kinetic energy is equal to the mean value of the potential energy.

Now according to the Law of Equipartition of Energy the energy associated with each degree of freedom is $RT/2$ ergs, and thus for the $\dot{x}$, $\dot{y}$, and $\dot{z}$ components of the oscillator the energy is $3RT/2$ ergs or $3T$ calories. Similarly for the x, y, and z components we have $3T$ calories or a total of $6T$ calories for the total kinetic and potential energies. The heat energy per degree, therefore (i.e. the specific heat), should be 6 calories. This is found to be, approximately, the case.

Now, if we take into consideration the effect of the free electrons and assume that their number is of the same order of magnitude as that of the atoms, the energy content of the metal should be increased by $\frac{3}{2}RT$ per gram-mol., corresponding to the three components of translatory motion of the electrons. Consequently, the specific heat of a metal should be $9R/2$ instead of $6R/2$ which, however, is not the case. This suggests that in some manner, as yet unexplained, the free electrons contribute little or nothing to the specific heat of a metal.

The Fermi-Dirac Theory

A way out of the foregoing difficulty may be found by assuming that the free electrons within a metal are quantized, i.e. that they are in stationary states. In these circumstances, providing they do not change from one state to another, their energy will be independent of the temperature of the metal. Now, whether an electron gas obeys gas or quantum laws, the energy of the electrons will be distributed according to some law, and it is this law and its consequences which it is desired to find. For a gas having a Maxwellian distribution of velocities the energy per molecule is $E = \frac{1}{2}mc^2$ and from (1–19), (1–26), and (1–34) we may find the number of molecules having energies lying within the range E and $E + dE$. We have

$$c^2 = 2E/m$$

differentiating, $\quad 2c\frac{dc}{dE} = \frac{2}{m}$

and

$$dc = \frac{dE}{mc} = \frac{dE}{\sqrt{2mE}} \quad . \quad . \quad . \quad (4\text{–}12)$$

From (1–26) and (1–34)

$$\sqrt{\frac{3kT}{m}} = \sqrt{\frac{3}{2}}a$$

hence

$$a = \sqrt{\frac{2kT}{m}} \quad . \quad . \quad . \quad . \quad . \quad (4\text{–}13)$$

Substituting from (4–10) and (4–11) into (1–19), we have

$$y = \frac{2n}{\sqrt{\pi}}\varepsilon^{-\frac{E}{kT}}\left(\frac{E}{kT}\right)^{\frac{1}{2}}\frac{dE}{kT} \quad . \quad . \quad (4\text{–}14)$$

$$= \frac{2n}{\sqrt{\pi}}\varepsilon^{-w}\,w^{\frac{1}{2}}dw \quad . \quad . \quad . \quad (4\text{–}15)$$

where y is the number of molecules with energies lying between E and $E + dE$ and $w = E/kT$.

The electron gas cannot conform to this law because, as we have already seen, its contribution to the specific heat of a metal is negligible. Furthermore, (4–14) indicates that the energy of a molecule can assume any value, whereas in the case of a gas governed by quantum conditions only certain energy states are possible. On the assumption that the electron gas is quantized, the conditions may be indicated by the energy-level diagram of Fig. 4–5, where the various horizontal lines represent different energy levels. They indicate that the free electrons within the metal possess a great variety of kinetic energies from zero up to the normal maximum level and that energies intermediate to these are impossible.

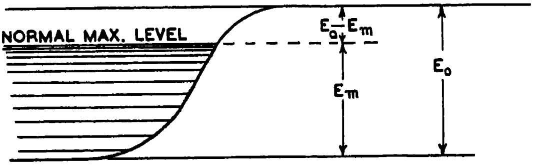


FIG. 4–5

In considering the quantum state of the free electrons within a metal, certain distinctions from previously considered quantum states must be noted. For example, in contrast with the linear oscillator, a free electron is in a region of practically uniform potential and thus its state is not defined by a position vector. Also such electrons are not undergoing periodic motion and hence are not governed by the phase integral $\int pdq$ where this is taken over a period of the system. These conditions are, of course, also true for the molecules of a perfect gas, but, in the case of a gas, the molecules may possess any energy and momentum values. In the case under consideration the state of each electron is solely represented by three momentum vectors p_x, p_y, p_z, which can be combined to form a single vector given by

$$p^2 = p_x^2 + p_y^2 + p_z^2$$

This vector will terminate in a volume-element, or cell, in the phase space of magnitude $dp_x dp_y dp_z dx dy dz$, or $dp_x dp_y dp_z dv$. From (1–75) we have

$$dp dq = h$$

and, as (in the present case) the motion of an electron along each of the three co-ordinates must be separately quantized, we may write

$$dp_x dx = h$$

$$dp_y dy = h$$

$$dp_z dz = h$$

Multiplying,

$$dp_x dp_y dp_z dx dy dz = h^3 \quad . \quad . \quad . \quad (4\text{–}16)$$

from which it is evident that the volume of a cell in the phase space is equal to h^3.

We shall now consider an electron gas within a metal with n electrons per unit volume. Let the momentum vectors of an electron be p_x, p_y, p_z, and consider the electrons with momenta between p and $p + dp$ which lie in a spherical shell of volume $4\pi p^2 dp$ in the momentum space. If there be A_s possible states for the electrons within this shell, it may be regarded as divided into A_s cells, or volume-elements. Now, according to Pauli's exclusion principle, a given combination of quantum numbers can occur but once in each atom. This principle may be extended to an electron gas, in which case it means that no two electrons may simultaneously possess the same amounts of action along their respective co-ordinates. Thus, each of the A_s cells can contain only one electron or none at all. From (4–16) the volume of a cell in the momentum space is

$$dp_x dp_y dp_z = h^3/dv$$

and hence the number of cells in the spherical shell $4\pi p^2 dp$ is

$$\frac{4\pi p^2 dp dv}{h^3} \quad . \quad . \quad . \quad . \quad . \quad (4\text{–}17)$$

Of the A_s cells, let A_{s0} contain no electrons, and A_{s1} one electron, so that $A_s = A_{s0} + A_{s1}$. Then, from (1–52) the number of different micro-states, or complexions, is

$$w_s = \frac{A_s!}{A_{s0}!\,A_{s1}!}$$

From the approximation formula of Stirling, $\log n! = n \log n - n$, we may write

$$\begin{aligned} \log w_s &= \log A_s! - \log A_{s0}!A_{s1}! \\ &= A_s \log A_s - A_s - A_{s0} \log A_{s0} + A_{s0} - A_{s1} \log A_{s1} + A_{s1} \\ &= A_s \log A_s - A_{s0} \log A_{s0} - A_{s1} \log A_{s1} \end{aligned}$$

The energy of an electron within the shell is $E_s = p^2/2m$ and thus the total energy within the shell is $A_{s1}E_s$. The entire volume of the momentum space may be regarded as divided into concentric shells $A_1, A_2, A_3, \ldots, A_s \ldots$, so that if w is the total number of different micro-states, then

$$\log w = \sum_s \log w_s = \Sigma(A_s \log A_s - A_{s0} \log A_{s0} - A_{s1} \log A_{s1})$$

$$\begin{aligned} &\text{Again} && \Sigma A_s = \Sigma(A_{s0} + A_{s1}) = \text{constant} \\ &\text{also} && E = \sum_s A_{s1}E_s \quad = \quad ,, \\ &\text{and} && n = \sum_s A_{s1} \quad = \quad ,, \end{aligned}$$

A state of equilibrium is attained when w is a maximum. Hence for equilibrium we must have $\log w$ a maximum, subject to the conditions expressed by the three equations immediately above. Regarding w_s, $A_{s1}E_s$ and A_{s1} as quasi-continuous functions of position and, applying the Calculus of Variations, we have, after multiplying the last three equations by the undetermined multipliers $-\alpha$, $-\beta$, and $-\gamma$, respectively,

$$\begin{aligned} (1 + \log A_{s1} + \alpha + \beta E_s + \gamma) &= 0 \\ (1 + \log A_{s0} + \alpha) &= 0 \end{aligned}$$

Hence

$$\begin{aligned} \log A_{s1} &= -(1 + \alpha + \beta E_s + \gamma) \\ \log A_{s0} &= -(1 + \alpha) \end{aligned}$$

from which

$$\log (A_{s0}/A_{s1}) = \gamma + \beta E_s$$

and

$$A_{s0} = A_{s1}\varepsilon^{\gamma + \beta E_s}$$

The number of electrons in the shell s is given by

$$n_s = A_{s1} = A_s \frac{A_{s1}}{A_{s0} + A_{s1}} = A_s \frac{1}{\frac{A_{s0}}{A_{s1}} + 1}$$

$$= A_s \frac{1}{\varepsilon^{\gamma + \beta E_s} + 1}$$

From (4–17),

$$A_s = \frac{4\pi p^2 dp dv}{h^3}$$

and, therefore,

$$n_s = \frac{4\pi p^2 dp dv}{h^3} \cdot \frac{1}{\varepsilon^{\gamma + \beta E_s} + 1}$$

Writing $\gamma = - \beta E_m$, the number of electrons per unit volume, the momenta of which lie between p and $p + dp$, is

$$n_1 = \frac{4\pi p^2 dp}{h^3} \cdot \frac{1}{\varepsilon^{\beta(E_s - E_m)} + 1} \quad . \quad . \quad (4\text{–}18)$$

So far we have not taken electron spin into consideration. The effect of this is to allow two electrons per cell, instead of one, because for every possible momentum value an electron has two possible states. Thus we must write (4–18)

$$n_1 = \frac{8\pi p^2 dp}{h^3} \cdot \frac{1}{\varepsilon^{\beta(E_s - E_m)} + 1} \quad . \quad . \quad (4\text{–}19)$$

The constant β may be shown to be equal to kT, where T is the absolute temperature and k is Boltzmann's constant.

From (4–17) the number of cells in the spherical shell $4\pi p^2 dp$ is $4\pi p^2 dp dv/h^3$, and from (4–17) and (4–19) the number of electrons per cell is

$$\frac{2}{\varepsilon^{(E_s - E_m)/kT} + 1} \quad . \quad . \quad . \quad (4\text{–}20)$$

When T is very small, or zero, (4–20) is equal to zero when $E_s > E_m$ and equal to 2 when $E_s < E_m$. This means that, at very low temperatures, each cell of the phase space for which the energy E_s, or $p^2/2m$, is less than E_m contains two electrons, while cells for which the energy is greater than E_m are empty. If we write $E_m = p_m{}^2/2m$, p_m is the maximum value the momentum may have for all the cells to be filled. The volume of a sphere in the momentum-space of radius p_m is $\frac{4}{3}\pi p_m{}^3$ and the number of cells per cubic centimetre within this

$$\tfrac{4}{3}\pi p_m{}^3/h^3$$

In each of these cells are two electrons, and hence

$$n = 2 \cdot \frac{\frac{4}{3}\pi p_m{}^3}{h^3}$$

where n is the number of electrons per cubic centimetre. Thus

$$E_m = \frac{p_m{}^2}{2m} = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3}$$

which shows that when $T = 0$ the maximum energy an electron can possess is proportional to $n^{\frac{2}{3}}$. Expressing E_m in electron-volts

$$E_m = \phi e = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3} \quad . \quad . \quad . \quad (4\text{–}21)$$

and

$$\phi = \frac{h^2}{2me}\left(\frac{3n}{8\pi}\right)^{2/3} \quad . \quad . \quad . \quad . \quad (4\text{–}22)$$

where ϕ is the potential difference necessary to give an electron energy E_m. On giving values to the various terms in (4–22), the following results are obtained—

n	ϕ
10^{21}	0·38
10^{22}	1·76
10^{23}	8·2
10^{24}	38
10^{25}	176

The number of free electrons per cubic centimetre in metals is of the order of 10^{23}, so that E_m is relatively large. A result of this is that the momentum distribution law expressed by (4–19) varies but little with temperature, even up to the melting point of a metal.

From the momentum distribution law of (4–19), velocity and energy distribution laws may be derived. The velocity of an electron is given by

$$p = mc$$

also

$$dp = mdc$$

Substituting in (4–19) we obtain

$$n_1 = \frac{8\pi m^3 c^2 dc}{h^3} \frac{1}{\varepsilon^{(E_s - E_m)/kT} + 1} \quad . \quad . \quad (4\text{–}23)$$

Again, the kinetic energy of an electron may be written

$$E = p^2/2m$$

from which

$$p^2 = 2mE$$

and

$$2pdp = 2mdE$$

or

$$dp = mdE/\sqrt{2mE}$$

Substituting in (4–19)

$$n_E = \frac{8\pi\sqrt{2}m^{3/2}\sqrt{E}dE}{h^3}\,\frac{1}{\varepsilon^{(E_s - E_m)/kT} + 1} \quad . \quad (4\text{–}24)$$

where n_E is the number of electrons per cubic centimetre having kinetic energies lying between E and $E + dE$.

For the case when $E_s < E_m$ and T is small, (4–24) becomes

$$n_E = \frac{8\pi\sqrt{2}m^{3/2}}{h^3}\sqrt{E}dE \quad . \quad . \quad . \quad (4\text{–}25)$$

from which it will be noted that the energy distribution follows a quadratic law. The energy distribution curve for tungsten is plotted for $T = 0$ in Fig. 4–6, the abrupt termination of the curve, of course, occurring for $E = E_m$.

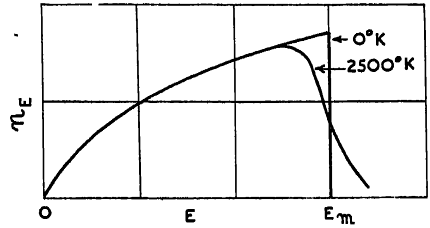


FIG. 4–6

As previously stated, variation of T has little effect on the distribution laws, and the energy distribution curve for tungsten at 2500° K is shown in Fig. 4–6, the appropriate equation in this case being (4–24). When $E = E_m$, n_E in both (4–24) and (4–25) have the same value. Hence, for all temperatures all curves must have the same ordinate for $E = E_m$, i.e.

$$\frac{8\pi\sqrt{2}m^{3/2}\sqrt{E}dE}{2h^3}$$

The areas under the curves are, of course, the number of electrons per cubic centimetre of metal. Thus, the areas of all curves are the same no matter what the temperature may be. Referring to the curve for 2500° K., it will be seen this curve approaches the abscissa asymptotically, indicating that a small number of electrons possess large energy values. It is these electrons which are involved in thermionic emission.

AVERAGE ELECTRON ENERGY

From (4–25) the total energy of the electrons per cubic centimetre is

$$\int_0^{E_m} n_E E = \frac{8\pi\sqrt{2}m^{3/2}}{h^3}\int_0^{E_m} E^{3/2}\,dE$$

$$= \frac{2}{5}\cdot\frac{8\pi\sqrt{2}m^{3/2}}{h^3}E_m^{5/2}$$

The number of electrons per cubic centimetre is

$$\frac{8\pi\sqrt{2}m^{3/2}}{h^3}\int_0^{E_m} E^{1/2}\,dE$$

$$= \frac{2}{3}\cdot\frac{8\pi\sqrt{2}m^{3/2}}{h^3}E_m^{3/2}$$

Hence the average energy per electron is

$$\frac{\frac{2}{5}\cdot\frac{8\pi\sqrt{2}m^{3/2}}{h^3}E_m^{5/2}}{\frac{2}{3}\cdot\frac{8\pi\sqrt{2}m^{3/2}}{h^3}E_m^{3/2}}$$

$$= \frac{3}{5}E_m$$

The subject of thermionic emission will now be considered in the light of the distribution law of (4–19). For this purpose the number of electrons having X-directed momenta between p_x and $p_x + dp_x$ must be found. Referring to Fig. 4–7, we have

$$p^2 = p_x^2 + p_y^2 + p_z^2$$

$$r^2 = p_y^2 + p_z^2$$

and the number of electrons having momenta between p_x and $p_x + dp_x$ will lie within an annulus of width dr and height dp_x. The volume of this is $2\pi r dr dp_x$, and as the number of electrons per cell in the momentum space is $2/h^3(\varepsilon^{(E_s - E_m)/kT} + 1)$, the number of electrons in the annular volume is given by

$$\frac{2\pi r dr dp_x \,.\, 2}{h^3\varepsilon^{(E_s - E_m)/kT} + 1}$$

and the total number of such electrons per cubic centimetre by

$$\frac{2dp_x}{h^3}\int_0^\infty \frac{2\pi r dr}{\varepsilon^{(E_s - E_m)/kT} + 1} \quad . \quad . \quad . \quad (4\text{–}26)$$

Now the electrons involved in thermionic emission are those the energies of which are in excess of E_m. Thus, if E_s is much larger than E_m, (4–26) may be written

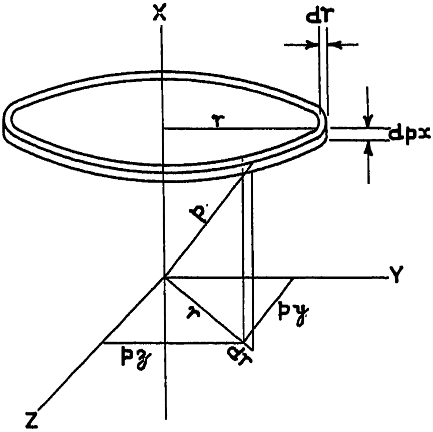


FIG. 4–7

$$\frac{2dp_x}{h^3}\int_0^\infty \varepsilon^{(E_m - E)/kT}\, 2\pi r dr = \frac{2dp_x}{h^3}\,\varepsilon^{E_m/kT}\int_0^\infty \varepsilon^{-E/kT}\, 2\pi r dr$$

Putting $E = (r^2 + p_x^2)/2m$, this becomes

$$\frac{4\pi}{h^3}\,\varepsilon^{E_m/kT}\,\varepsilon^{-p_x^2/2mkT}\,dp_x\int_0^\infty \varepsilon^{-r^2/2mkT}\, r dr$$

Integrating [see (1–16)], we have

$$\frac{4\pi}{h^3}\,\varepsilon^{E_m/kT}\, mkT\varepsilon^{-p_x^2/2mkT}\,dp_x$$

which is the number of electrons per cubic centimetre with momenta lying between p_x and $p_x + dp_x$. The number of electrons per second passing through 1 cm.^{2} perpendicular to the x direction with x-directed momentum greater than p_0 is, therefore,

$$4\pi\varepsilon^{E_m/kT}\frac{mkT}{h^3}\int_{p_0}^{\infty}\varepsilon^{-p_x^2/2mkT}\frac{p_x}{m}\,dp_x$$

If all these escape from the metal, the thermionic current density is

$$J = 4\pi\varepsilon^{E_m/kT}\frac{ekT}{h^3}\int_{p_0}^{\infty}\varepsilon^{-p_x^2/2mkT}p_x dp_x$$

$$= \frac{4\pi emk^2T^2}{h^3}\varepsilon^{(E_0-E_m)/kT} \quad . \quad . \quad . \quad (4\text{–}27)$$

where $E_0 = p_0^2/2m$. It will be noted that this result agrees with those found from classical theories provided $\phi e = (E_0 - E_m)$. Thus, the minimum energy for escape is not ϕe but $\phi e + E_m$. As we have already seen, ϕ may have any value up to about 7 volts, and, as E_m is about 10 electron-volts, E_0 must lie between 10 and 17 electron-volts.

Returning to the energy-level diagram of Fig. 4–5, this may now be clarified. What was formerly termed the normal maximum level is now seen to correspond to the maximum energy which an electron can possess at low temperatures. The quantity E_m has been termed the "inner work function" and E_0 the "outer" or "gross work function." The difference in the two quantities is termed the "net work function," or, more commonly, the work function. The number of electrons per level has been shown to be 2, and because the number of electrons the energy of which lies between E and $E + dE \propto \sqrt{E}$, it is evident that the levels must lie closer together as E increases.

VELOCITY DISTRIBUTION OF EMITTED ELECTRONS

Although the energy and velocity distributions of the electrons within a metal follow the Fermi-Dirac Law, electrons emitted thermionically follow Maxwell's Law. In order to study the conditions prevailing among the electrons emitted from a hot body, we shall consider a heated wire surrounded by a concentric cylinder. The wire is assumed to be straight and of small diameter, and the number of electrons emitted from it sufficiently small for space-charge effects to be negligible. The electrons are emitted with a certain velocity which can be resolved into two components, one parallel to the wire and the other radial. As the former contributes nothing to the thermionic current it may be neglected. In order that an electron may cross the gap from the wire to the cylinder it must possess a certain amount of energy. The energy associated with the radial velocity component is $\frac{1}{2}mv_r^2$. Now if a voltage is applied to wire and cylinder so that the cylinder is negative, this will constitute a retarding potential which will tend to prevent electrons from reaching the cylinder. In order that an electron may cross the gap it is evident that

$$\tfrac{1}{2}mv_r^2 \gtrsim -eV \qquad . \quad . \quad . \quad (4\text{–}28)$$

where V is the potential difference between wire and cylinder. $-eV$ is, of course, the energy necessary to make the crossing.

Now the probability of an electron having energy lying between E and $E + dE$ is $f(E)dE$, and hence the number of electrons having energies between these limits is

$$dn = nf(E)dE$$

where n is the number of electrons emitted per second. The saturation current is $ne = i_0$ and hence the current is given by

$$i = i_0\int_{-eV}^{\infty} f(E)dE$$

Substituting $-eV$ for E, we have

$$\frac{di}{dV} = -i_0ef(-eV)$$

and if i is found for different values of V, di/dV may be determined, and hence $f(E)$. Experiments give $i/i_0 = \varepsilon^{-eV/kT}$ so that

$$\frac{di}{dV} = -\frac{i_0e}{kT}\varepsilon^{-eV/kT} = -i_0ef(-eV)$$

from which $f(E) = \dfrac{1}{kT}\varepsilon^{-E/kT}$

Thus the number of electrons with radial-associated energy lying between E and $E + dE$ is

$$dn = \frac{n}{kT}\varepsilon^{-E/kT}dE \qquad . \quad . \quad . \quad (4\text{–}29)$$

and the fraction of electrons having this energy is

$$\frac{dn}{n} = \frac{1}{kT}\varepsilon^{-E/kT}dE \qquad . \quad . \quad . \quad (4\text{-}30)$$

Now according to Maxwell's Velocity Distribution Law the number of molecules per cubic centimetre having velocities between u and $u + du$ is

$$\frac{n}{a\sqrt{\pi}}\varepsilon^{-\frac{u^2}{a^2}}du$$

and the number of these passing through 1 cm.^{2} perpendicular to u per second is

$$\frac{n}{a\sqrt{\pi}}\varepsilon^{-\frac{u^2}{a^2}}udu\ . \qquad . \qquad . \qquad . \quad (4\text{–}31)$$

The total number passing per second is

$$\frac{n}{a\sqrt{\pi}}\int_0^{\infty}\varepsilon^{-\frac{u^2}{a^2}}udu$$

$$= \frac{na}{2\sqrt{\pi}} \qquad . \qquad . \qquad . \qquad . \quad (4\text{–}32)$$

and thus the fraction passing per second is

$$\frac{2}{a^2}\varepsilon^{-\frac{u^2}{a^2}}udu \quad . \qquad . \qquad . \qquad . \quad (4\text{–}33)$$

Putting $\qquad \frac{1}{2}mu^2 = E,\ u = \sqrt{\frac{2E}{m}},$

$$du = \frac{dE}{mu} = \frac{dE}{\sqrt{2mE}}$$

Also $\qquad \frac{1}{a^2} = \frac{m}{2kT}\ .$

Substituting in (4–33), the fraction is

$$\frac{1}{kT}\varepsilon^{-E/kT}dE$$

which is identical with the result given by (4–30). Hence the energy and velocity distributions of the electrons outside the metal is Maxwellian.

Emission Constants

It is customary to write (4–7) in the form

$$J = AT^2\varepsilon^{-\frac{b}{T}}\ . \qquad . \qquad . \qquad . \quad (4\text{–}34)$$

where A and b are constants. Actually, however, b may vary slightly with temperature. As the expression for A contains only universal constants, it follows that it should be the same for all metals. Substitution of the numerical values gives

$$A = 120{\cdot}4 \text{ amp./cm.}^2 \text{ per deg.}^2 \text{ K.}$$

Also $$b = (E_0 - E_m)/k = \phi e/k = 11{,}600\,\phi \text{ deg. K.}$$

In practice, A is found to have a value about one-half of that derived above, i.e. 60 amp./cm.^{2} No satisfactory explanation has yet been given of this, but it has been suggested that a reflexion coefficient of 0·5 would account for the discrepancy. Some values of emission constants for pure metals are given by Table 4–1.

TABLE 4–1

METAL	A	b	ϕ	MELTING POINT ° K.
Calcium	60·2	26,000	2·24	1,083
Cesium	16·2	21,000	1·81	299
Carbon	60·2	46,500	3·82	3,773
Molybdenum	60·2	51,500	4·43	2,893
Platinum	60·2	59,000	5·08	2,028
Nickel	26·8	32,100	2·77	1,725
Tantalum	60·2	47,200	4·06	3,123
Tungsten	60·2	52,400	4·52	3,643
Thorium	60·2	38,900	3·35	2,118

Figures such as those given are obtained with metals of a high degree of purity, outgassed, in the best vacuum obtainable. In spite of this, however, considerable variations are shown between various experiments and experimenters.

DETERMINATION OF EMISSION CONSTANTS

In order to determine emission constants the arrangement of Fig. 4–8 may be employed. This consists of a filamentary cathode surrounded by a concentric cylindrical anode with guard rings to eliminate end effects. This arrangement is placed in a high vacuum and the temperature of the filament measured by means of an optical pyrometer sighted on the filament through the hole H. The anode-cathode potential difference is raised to such a value that all the electrons are drawn to the cathode and the saturation current results. Hence the value of this current may be measured for various values of cathode temperature T. Dividing (4–34) by T^2, and taking the logarithm of both sides

$$\log_{10}\frac{J}{T^2} = -\,0{\cdot}434b\,\frac{1}{T} + \log_{10} A \qquad . \qquad . \quad (4\text{–}35)$$

From this it follows that a plot of experimental values should give a straight line when arranged as indicated by (4–35). The slope of this line is given by $-\,0{\cdot}434b$, from which b may be derived. The intercept of this line on the ordinate gives $\log_{10} A$ and consequently A. This value can be obtained only by producing the line, for interception occurs for $T = \infty$. A result obtained with a tungsten cathode is shown by Fig. 4–9.

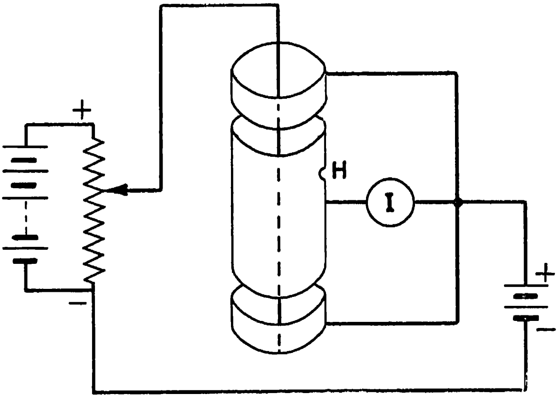


Fig. 4–8

In practice, considerable difficulties are experienced in obtaining accurate and consistent values of emission constants. As an indication of the relative importance of the measurements, the ratio of the percentage change in current for a given percentage change in temperature will be found. Differentiating (4–34)

$$\frac{dJ}{dT} = A\varepsilon^{-\frac{b}{T}}(b + 2T)$$

Multiplying by T/J we have

$$\frac{dJ/J}{dT/T} = \frac{b}{T} + 2$$

If a practical value of about 23 is taken for b/T, it is evident that the measurement of temperature should be 25 times as accurate as that of the measurement of current in order that the derived value of A shall be as precise as the measurement of J.

It has already been stated that a high vacuum must be employed in the determination of emission constants. In addition, all absorbed and occluded gases must be removed, this necessitating baking of the glass envelope and outgassing of the electrodes, etc. In practice, although the range of temperature over which investigation may be made is not large, the thermionic current range is relatively enormous. Thus, for tungsten, if the temperature range is from 1000° K. to 3000° K., the current ratio corresponding to these limits is of the order of 10^{12}.

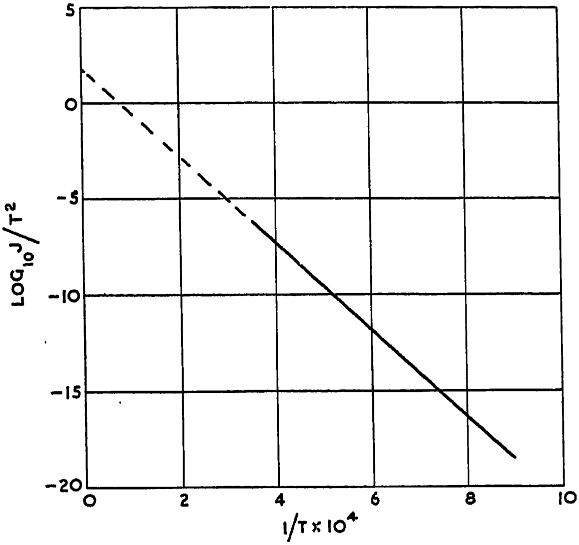


FIG. 4–9

CATHODE TEMPERATURE

When the cathode is not visible it is, of course, impossible to employ an optical pyrometer for the determination of its temperature. In such circumstances it is possible to obtain the temperature with the aid of the Stefan-Boltzmann Law. When the cathode is in a state of temperature equilibrium, the rate at which energy is supplied to it is equal to the rate at which it is parting with energy in the form of heat. As the cathode is situated in a vacuum, convection losses are negligible and the bulk of the loss occurs through radiation. A small amount of heat is lost by conduction through the cathode supports and a smaller amount by the emission itself. The latter is due to electron evaporation and corresponds to the cooling effect experienced with an evaporating liquid.* In order to form an idea of the heat-loss due to emission, the energy carried off from the cathode by the electrons must be found. As we have already seen, the energy and velocity distribution of the electrons passing through the surface follows the Maxwellian Distribution Law, and hence we must find the average energy of an electron outside the cathode in accordance with this law. If E_a is this energy, then the total energy possessed by an electron before escape is $E_0 + E_a$, this being reduced to E_a after escape. As before, E_0 is the outer or gross work function of the metal forming the cathode.

Now E_a will consist of three parts: the x-, y-, and z-associated energies of an electron. As it is the x-associated energy which is responsible for electron-escape, we shall consider this first. According to (4–32) the number of electrons per unit volume which have x-directed motion and pass through each square centimetre per second is

$$\frac{an}{2\sqrt{\pi}} \quad . \quad . \quad . \quad . \quad (4\text{–}36)$$

the result being for positively directed motion only. Transforming (4–31) into an energy expression, as on page 161, we have

$$\frac{n}{ma\sqrt{\pi}}\,\varepsilon^{-\frac{2E}{ma^2}}\,dE$$

which gives the number of electrons having energies lying between E and $E + dE$ passing through each square centimetre per second. The energy carried by these is

$$\frac{n}{ma\sqrt{\pi}}\,\varepsilon^{-\frac{2E}{ma^2}}\,EdE$$

and hence the total energy of all such electrons is

$$\frac{n}{ma\sqrt{\pi}}\int_0^\infty \varepsilon^{-\frac{2E}{ma^2}}\,EdE$$

$$= \frac{nm^2a^4}{4ma\sqrt{\pi}}$$

---

* The principle of the wet-bulb thermometer, of course, depends on this phenomenon.

Dividing by (4–36), the average energy per electron is

$$\frac{nm^2a^4 2\sqrt{\pi}}{4ma\sqrt{\pi}an}$$
$$= \frac{ma^2}{2}$$

But $a^2 = 2kT/m$, which finally gives the x-associated energy as kT per electron. Now the y- and z-associated energies of an electron do not contribute to escape and thus these energies are the same outside the cathode as inside. They are governed by the Law of Equipartition of Energy, according to which each electron has an average energy equal to $kT/2$. Hence, the total average energy with which an electron leaves the cathode surface is

$$E_a = kT + kT/2 + kT/2$$
$$= 2kT$$

Of the minimum energy needed for electron-escape, i.e. E_0, we have already seen that E_m is always present. Thus, the minimum energy which must be supplied by the cathode heating source for escape is $E_0 - E_m = \phi e$. As, in addition to this, the electron escapes with an average energy $2kT$, the total energy which must be supplied is $(\phi e + 2kT)$. If $e\phi_r$ is the electron-volt equivalent of kT, then the power loss due to electron emission is $J(\phi + 2\phi_r)$ watts/cm.^{2} of cathode surface. It will be immediately evident that this quantity is virtually negligible compared with radiation losses.

CATHODE TEMPERATURE DETERMINATION

Assuming that there is little heat shielding from neighbouring surfaces, the rate of energy radiation from a heated cathode may be expressed by the Stefan-Boltzmann Law

$$P = \sigma\gamma T^4$$

where σ has the value given on page 31 and γ is the emissivity coefficient of the cathode surface. This coefficient is always less than unity and is the ratio of the power radiated from the surface to that radiated by a black body at the same temperature. It is also the fraction of incident radiation absorbed by a grey surface. Hence we must regard a cathode as a grey-body radiator. In practice γ is not strictly constant, but varies slightly with temperature. If P above is expressed in watts, then the right-hand member must be divided by 10^7. This gives

$$P = 5{\cdot}75 \times 10^{-12}\gamma T^4 \text{ watts/cm.}^2 \quad . \quad . \quad (4{\cdot}37)$$

In the event of the cathode being heat-shielded, heat will be reflected and the temperature of the surroundings may be appreciable. In these circumstances (4–37) must be replaced by

$$P = 5{\cdot}75 \times 10^{-12}\gamma(T^4 - T_0{}^4) \text{ watts/cm.}^2$$

where T_0 is the surrounding temperature. However (apart from special heat-shielded cathodes described in Chapter IX), $T_0{}^4$ is usually negligible compared with T^4, and hence (4–37) may be employed.

Taking the logs of both sides of (4–37), we have

$$\log_{10} P = \log_{10} (5{\cdot}75 \times 10^{-12}\gamma) + 4 \log_{10} T$$

which, if γ is constant, gives a straight line with a slope equal to 4. In practice the line is slightly curved, indicating that γ is not constant. Hence γ must be experimentally determined for various temperatures. For example, for what is known as Davisson's combined-type coated cathode, (4–37) must be written

$$P = 5{\cdot}75 \times 10^{-12} [0{\cdot}4 + (2{\cdot}5 \times 10^{-4}T)]T^4$$

where $\gamma = [0{\cdot}4 \times (2{\cdot}5 \times 10^{-4}T)]$

If the cathode input watts are measured, then

$$P = \frac{W}{S}$$

where W is the input watts and S the cathode area. The cathode temperature is then determined from (4–37) by writing

$$T^4 = \frac{P}{5{\cdot}75 \times 10^{-12}\gamma}$$

or

$$T = \frac{P^{\frac{1}{4}}}{1{\cdot}55 \times 10^{-3}\gamma^{\frac{1}{4}}} \quad . \quad . \quad . \quad (4–38)$$

Again, if the thermionic saturation current density is known, T may be determined. From (4–35)

$$\log_{10} J - 2 \log_{10} T = \log_{10} A - 0{\cdot}434 \frac{b}{T} \quad . \quad (4–39)$$

Rearranging, there results the transcendental equation

$$T = \frac{0{\cdot}434b}{2 \log_{10} T - \log_{10} J/A}$$

which may be solved by successive approximations.

If the value of T given by (4–38) is substituted in (4–39), J may be determined for any value of P. Thus, we have

$$\log_{10} J - 2\log_{10}\frac{P^{\frac{1}{4}}}{1{\cdot}55 \times 10^{-3}\gamma^{\frac{1}{4}}} = \log_{10} A - 0{\cdot}434\,\frac{b \times 1{\cdot}55 \times 10^{-3}\gamma^{\frac{1}{4}}}{P^{\frac{1}{4}}}$$

A further method of determining the cathode temperature is in terms of the hot and cold resistances of the cathode. The resistance of most conductors may be expressed as

$$R = R_0(1 + \alpha T + \beta T^2) \qquad . \quad . \quad (4\text{–}40)$$

where R_0 is the resistance at 0° C., R that at temperature T, and α and β are constants. Solving for T,

$$T = \frac{-\alpha + \sqrt{\alpha^2 - 4\beta\left(1 - \frac{R}{R_0}\right)}}{2\beta}$$

the result, in this instance, being in degrees centigrade. If, as is more customary, the initial resistance is taken at some other temperature than 0° C., say, T_1, then calling the resistance at T_1, R_1, (4–40) may be written

$$R = R_1[1 + \alpha_1(T - T_1) + \beta_1(T - T_1)^2]$$

which leads to

$$T - T_1 = \frac{-\alpha_1 + \sqrt{\alpha_1{}^2 - 4\beta_1\left(1 - \frac{R}{R_1}\right)}}{2\beta_1}$$

If T and T_1 are expressed in ° K., and T_1 is taken as 293°, then

$$T = \frac{-\alpha_1 + \sqrt{\alpha_1{}^2 - 4\beta_1\left(1 - \frac{R}{R_1}\right)}}{2\beta_1} + 293$$

Fig. 4–10 gives R/R_1 plotted against T,* the values of α_1 and β_1 being respectively $4{\cdot}7 \times 10^{-3}$ and $5{\cdot}2 \times 10^{-7}$ for the case shown (tungsten).

EMISSION EFFICIENCY

The electron-emitting efficiency of a cathode is expressed as

$$\frac{\text{Thermionic Current Obtainable}}{\text{Cathode Heating Power}} \qquad . \quad . \quad (4\text{–}41)$$

and it is evidently desirable that this should be as high as possible, consistent with other requirements. In general, a cathode with a low value of *b*, i.e. a low work function, provides a relatively large emission at a low temperature. However, the value of *b* is by no means the only criterion of the quality of an emitter. As the heating power and temperature of an emitter are raised, (4–41) rapidly increases, with the result that the most suitable operating temperature of a cathode is the highest possible temperature compatible with a rate of evaporation of cathode material that results in a reasonable cathode life. Tungsten, for example, because of its low vapour pressure and high melting point, can be operated at a sufficiently high temperature to give a greater emission than any other pure metal, although, as will be seen from Table 4–1, its value of *b* is higher than that for most other metals. Thus many materials, although having a low value of *b*, are unsuitable as emitters, because rapid evaporation prevents their being raised to a temperature at which efficient emission occurs.

---

* °K.

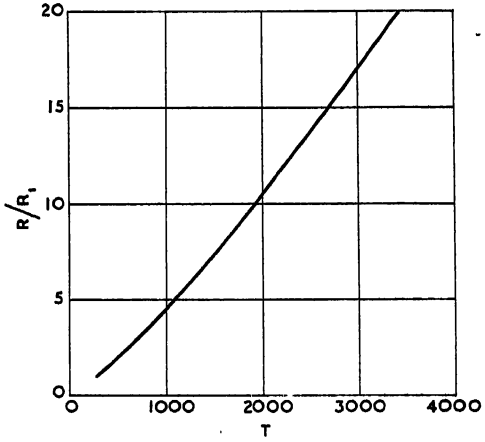


FIG. 4–10

This subject will be further discussed in Chapter VIII.

Contact Potentials

If two dissimilar metals, i.e. metals with different work functions, are placed in contact at one point only, a potential difference may be observed at the free ends. (See Fig. 4–11.) This potential difference, however, cannot be detected with an ordinary electromagnetic voltmeter for reasons which will be apparent later. Measurement of this potential difference shows that it corresponds to the difference of the work functions of the two metals. When the two metals are joined it may be taken as an experimental fact that the work done in passing electrons from one to the other is very small. This means that the potential barriers of the metals disappear at the point of contact and that electrons flow freely from one metal to the other by virtue of their kinetic energies. Actually, a small amount of work is done (should electrons pass from one metal to the other) corresponding to a potential difference of the order of millivolts, known as the Peltier effect. Ignoring this, however, it may be taken that on first connecting two metals at one point only electrons initially flow from one to the other until their potential is the same. This means that the metal receiving electrons acquires a negative charge which reduces its potential and sets up a field which finally reduces to zero the tendency of electrons to flow in its direction. When a state of equilibrium is attained, as many electrons flow in one direction as in the other during a given time. In the equilibrium state, and at low temperatures, it is evident that the E_m energy levels of the two metals are aligned at the point of contact. Now, if an electron is taken round the circuit, i.e. through each metal and across the air space between the free ends in turn, the work done will be zero, i.e.

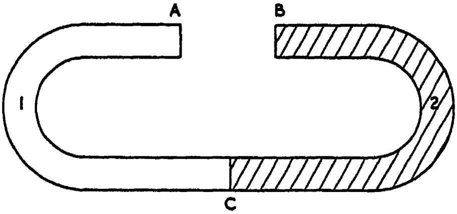


FIG. 4–11

$$e\phi_2 - e\phi_1 + e\phi_c = 0 \qquad . \quad . \quad . \quad (4\text{–}42)$$

where ϕ_2 and ϕ_1 are, respectively, the voltages corresponding to the work functions of the two metals and ϕ_c is the contact potential difference, i.e. the potential difference between the free faces of the metals. Dividing (4–42) by e, we have

$$\phi_c = \phi_2 - \phi_1 \text{ volts}$$

It has previously been stated that when the metals are placed in contact, electrons initially flow from one to the other. The direction of flow is from the metal of lower work function to that of higher work function and, consequently, it is the latter which acquires a negative charge. After an equilibrium state is attained, as many electrons pass from one side as from the other in a given time. Those that pass to the metal of higher work function, i.e. the one with a negative charge, must have work done on them to urge them towards this charge. Hence, in this respect their potential energy increases. However, as previously stated, joining the metals aligns their energy levels and thus there is no change in the energy of an electron as it passes from one metal to the other. This means that as an electron passes across the junction to the metal of higher work function, its kinetic energy falls, and vice versa when it passes in the opposite direction. Thus, if $p_{x1}^2/2m$ is the x-associated kinetic energy of an electron in the metal of lower work junction and $p_{x2}^2/2m$ that of an electron in the other metal, then,

$$p_{x1}^2/2m = p_{x2}^2/2m + e\phi' \quad . \quad . \quad . \quad (4\text{–}43)$$

where $e\phi'$ is the difference of the kinetic energies in the two metals. Now, from page 169 we have

$$\frac{4\pi}{h^3}\,\varepsilon^{E_{m1}/kT}\,mkT\varepsilon^{-p_{x1}^2/2mkT}\,dp_{x1}$$

$$\frac{4\pi}{h^3}\,\varepsilon^{E_{m2}/kT}\,mkT\varepsilon^{-p_{x2}^2/2mkT}\,dp_{x2}$$

which give the numbers of electrons within ranges p_{x1} and $p_{x1} + dp_{x1}$ and also p_{x2} and $p_{x2} + dp_{x2}$ in each metal. The numbers of electrons per second passing through unit area of the junction of the two metals is, for the ranges given,

$$\frac{4\pi}{h^3}\,kT\varepsilon^{(E_{m1} - p_{x1}^2/2m)/kT}\,p_{x1}dp_{x1} \quad . \quad . \quad (4\text{–}44)$$

and

$$\frac{4\pi}{h^3}\,kT\varepsilon^{(E_{m2} - p_{x2}^2/2m)/kT}\,p_{x2}dp_{x2} \quad . \quad . \quad (4\text{–}45)$$

When a state of equilibrium exists (4–44) and (4–45) are equal, also, from (4–43), $p_{x1}dp_{x1} = p_{x2}dp_{x2}$. Substituting for $p_{x1}^2/2m$ in (4–44), from (4–43), (4–44) and (4–45) can only be equal for

$$E_{m1} - p_{x2}^2/2m - e\phi' = E_{m2} - p_{x2}^2/2m$$

or

$$e\phi' = E_{m1} - E_{m2} \quad . \quad . \quad (4\text{–}46)$$

Thus an electron, in passing from the metal of lower work function to that of higher, experiences a change of kinetic energy equal to the difference of the normal maximum energy levels of the two metals. Or, we may say that the change is equal to the difference in the respective inner work functions of the metals. It is, of course, possible to express (4–46) in terms of the electron densities of the two metals. From (4–21) we have

$$e\phi' = \frac{h^2}{2m}\left(\frac{3}{8\pi}\right)^{\frac{2}{3}}\left(n_1^{\frac{2}{3}} - n_2^{\frac{2}{3}}\right)$$

where n_1 and n_2 are respectively the numbers of electrons per cm.^{3} in the two metals.

From the foregoing results it is now possible to draw the potential-energy diagram for two metals in contact. Referring to Fig. 4–12, it will be noted that the potential barriers have been drawn as vertical lines instead of curves as shown by Fig. 4–5. This may be taken as permissible, as the distance between the free surfaces is large compared with the atomic dimensions over which the curve of Fig. 4–5 holds. As the maximum-energy levels are aligned at the point of contact of the metals, it is evident that the difference in the zero-energy levels is as shown. As previously stated, the energy difference between the free surfaces is equal to the difference of the work functions of the metals.

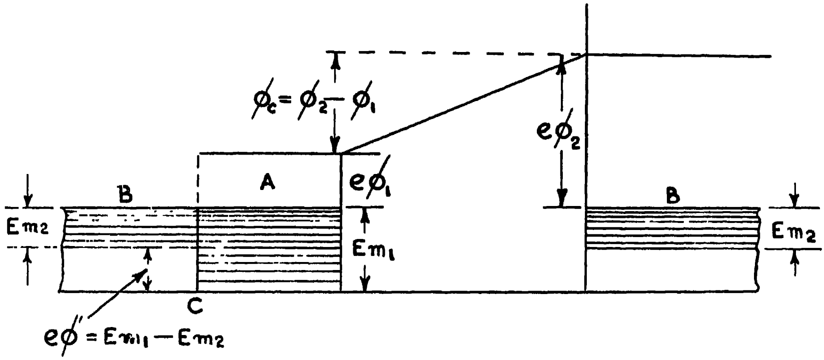


Fig. 4–12

In the event of a third metal being inserted at the junction C, this has no effect on the conditions in the space between the metals A and B. On the diagram it is only necessary to insert the energy levels of the third metal in such a manner that its normal maximum level coincides with the other two. It follows that if any number of dissimilar metals are connected in series, the contact potential of the combination corresponds to the difference of the work functions of the pair of metals which terminate the combination.

Should a pair of metals be connected via a source of potential, such as a battery, then the potential difference between the free faces is equal to the algebraical sum of the potential of the battery and that corresponding to the difference of the work functions of the metals.

Field or Auto-Electronic Emission

Although the previously given equations for the thermionic saturation current indicate a definite limiting current corresponding to a given cathode temperature, it is found in practice that the thermionic current does not truly saturate, but continues to increase slowly as long as the anode-cathode potential is increased. An explanation of this phenomenon was first given by Schottky, the phenomenon now being known as the Schottky effect. According to electrostatic theory, the attraction between a negatively-charged body and an equipotential surface is due to the positive charge induced by the body on the surface; alternatively, it is said that the body is attracted by its "image" in the surface, the attractive force being termed the *image force*. According to Schottky, the potential barrier, or work function, is solely due to this image force, and he calculated the modifying effect on the work function of an electrostatic field at a metal surface.

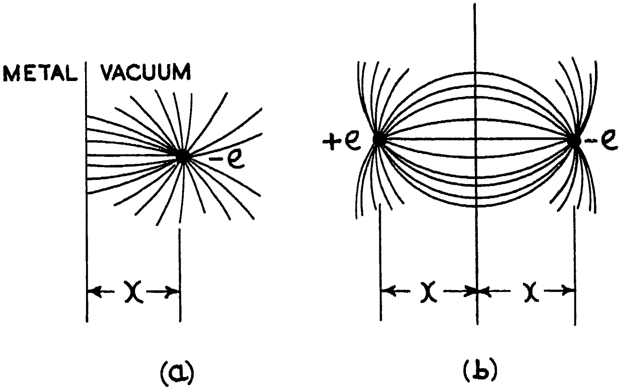


FIG. 4–13

Assuming the surface of a thermionic emitter to be a perfect plane, the conditions due to an electron at a distance x from the surface may be represented by Fig. 4–13, where (*a*) shows the actual condition and (*b*) that when the image hypothesis is employed. The force between plane and electron is

$$f = \frac{e^2}{(2x)^2 K} = \frac{e^2}{4Kx^2} \quad . \quad . \quad . \quad (4\text{–}47)$$

where K is the specific inductive capacity of the space and may be taken as unity. This equation is not applicable at distances very close to the surface, the law changing as the crystal lattice is approached. The actual value of the force for very small values of x depends on whether the electron approaches an atom or the space between the lattice points. Should an electric field exist at the surface of the emitter, the force on the electron will be modified from that given by (4–47). Close to the surface the image force will predominate, while at large values of x the field force, $-eX$, will be the dominating influence. At some distance x_m the two forces are equal and opposite, and at this position the electron experiences no force. Beyond x_m the field force tends to remove the electron from the influence of the emitter, and the electron will escape, providing it leaves the surface with sufficient kinetic energy to reach x_m. As it is unlikely that fields can be produced of such strength as to render x_m less than that at which (4–47) is applicable, we may write

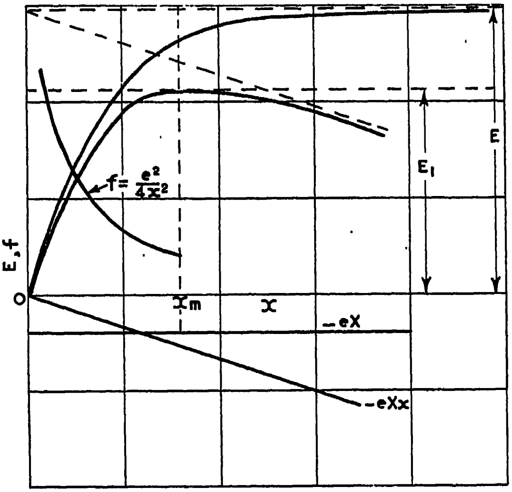


Fig. 4–14

$$\frac{e^2}{4x_m^2} = eX$$

and

$$x_m = \tfrac{1}{2}\sqrt{\frac{e}{X}} \quad . \quad . \quad . \quad . \quad (4\text{–}48)$$

Reference to Fig. 4–14 will show the image force, the field force, the potential energy in the absence of a field, and that in the presence of a field, all as functions of x. The reduction in the energy needed to remove the electron to infinity, i.e. $E - E_1$, corresponds to a lowering of the potential energy barrier or work function. This reduction may be considered in two parts: that needed to translate the electron from the metal to x_m, and that to remove the electron from x_m to infinity. The reduction in work, due to the field, for the first part is

$$eXx_m$$

Regarding the second part, when the electron reaches x_m, in the presence of an applied field, it is free to escape. Hence the reduction in work is equal to that needed to remove the electron from x_m to infinity in the absence of the field. This work is

$$\int_{x_m}^{\infty} \frac{e^2}{4x^2}\,dx = \frac{e^2}{4x_m}$$

and thus we have

$$E - E_1 = eXx_m + \frac{e^2}{4x_m}$$

where E is the work in the absence of a field and E_1 that in its presence. Substituting for x_m from (4–48)

$$\begin{aligned} E - E_1 &= \frac{1}{2}eX\sqrt{\frac{e}{X}} + \frac{1}{2}e^2\sqrt{\frac{X}{e}} \\ &= e\sqrt{eX} \end{aligned}$$

Expressing this in terms of the change in the work function

$$dE = E - E_1 = e\sqrt{eX}$$

and

$$E_1 = E - e\sqrt{eX} \quad . \quad . \quad . \quad . \quad (4\text{–}49)$$

where E_1 is the work function in the presence of the field. It will be noted that the reduction in the work junction is directly proportional to the square root of the field strength.

Considering now the effect on the emission current, let J_0 be the current density in the absence of a field and J that in its presence. Then, from (4–34)

$$J_0 = AT^2\varepsilon^{-\frac{E}{kT}}$$

and

$$J = AT^2\varepsilon^{-\frac{E_1}{kT}}$$

From (4–49) $$J = AT^2\varepsilon^{-\frac{E}{kT}}\varepsilon^{\frac{e\sqrt{eX}}{kT}}$$

or $$J = J_0\varepsilon^{\frac{e\sqrt{eX}}{kT}}$$

Substituting for e and k and expressing X in volts per centimetre, we have

$$J = J_0\varepsilon^{4\cdot4\sqrt{X}/T} \quad . \quad . \quad . \quad . \quad (4\text{–}50)$$

Multiplying both sides of (4–50) by the cathode area gives the emission current, viz.

$$I = I_0\varepsilon^{4\cdot4\sqrt{X}/T} \quad . \quad . \quad . \quad . \quad (4\text{–}51)$$

Now X is proportional to the anode voltage, with the result that, in practice, the relation between I and anode voltage may be shown in the manner indicated by Fig. 4–15. In order to determine the true saturation current as expressed by (4–51), we have

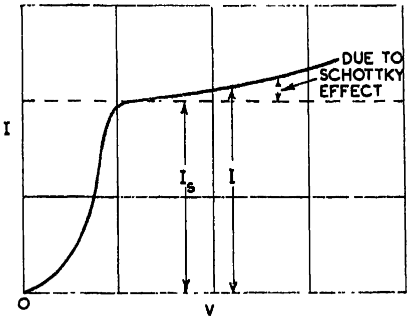


FIG. 4-15

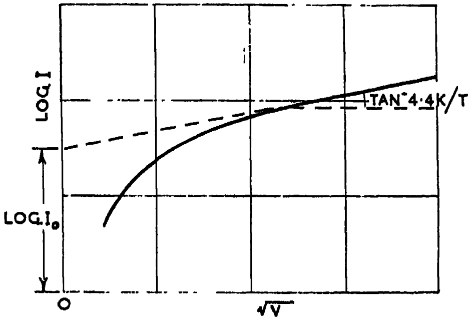


FIG. 4–16

$$\log I = \log I_0 + 4\cdot4\sqrt{X}/T$$

or $$\log I = \log I_0 + 4\cdot4K\sqrt{V}/T$$

where V is the anode voltage and K is a constant. At some value of V, I_0 becomes constant and thereafter log I is a straight line the slope of which is $4\cdot4K/T$. The intercept of this line on the ordinate then gives log I_0 in the manner shown by Fig. 4–16.

Photo-electricity

Reference has already been made to the photo-electric effect in Chapter II and its influence in reducing the statistical time lag of a spark-gap. In general it is found that the influence of radiation on metals is to liberate electrons from the latter. That this is so was originally demonstrated by J. J. Thomson, who measured e/m for the particles liberated and found it to be that for the electron. The kinetic energy of the emitted electrons can be determined by the potential difference required to arrest them.

Suppose a metal plate at a potential V is emitting electrons under the influence of radiation. An electron will experience a force attracting it back to the plate, and by the time it reaches a situation of zero potential it will have lost energy equal to Ve. If its original energy at the plate were E, then the energy is now $E - Ve$. If this quantity is equal to or greater than zero it will escape being drawn back to the plate. If v is the initial velocity of the electron, then, for escape,

$$\tfrac{1}{2}mv^2 \geq Ve$$

and

$$v^2 \geq 2V\frac{e}{m}$$

If a plate is enclosed in a highly-evacuated glass vessel with another conductor at zero potential, illumination of the plate, with, say, ultra-violet radiation, will cause the emission of electrons. The plate will slowly acquire a positive potential which will increase until it is just sufficient to prevent the escape of the fastest electrons. This maximum potential is a measure of the maximum kinetic energy of an electron. It is found to be of the order of 1 volt, and thus the energy of escape is

$$Ve = \frac{1}{300} \times 4{\cdot}774 \times 10^{-10}$$

$$\doteqdot 1{\cdot}591 \times 10^{-12} \text{ ergs.}$$

i.e. 1 electron-volt. It is found that the velocity and energy of the emitted electrons are independent of the intensity of the incident radiation, but depend on the frequency of the latter.

MILLIKAN'S EXPERIMENTS

There have been many experimenters in the field of photoelectricity, but probably the best results have been obtained by Millikan, whose experiments alone will be described here. It is found that alkali metals emit most readily and that oxide films, etc., on the metal surface modify and complicate results. In view of this, Millikan's experiments were carried out on a block of alkali metal supported within a highly-evacuated glass vessel. In order to obtain a perfectly clean surface the metal was scraped inside the vacuum by a cutter operated from outside by means of an electromagnet. The clean surface of the alkali was situated in front of an oxidized copper-gauze cylinder, the potential difference between the two being adjusted until no electrons could escape from the former. The purpose of employing copper oxide for one electrode is that the photo-electric effect occurs at lower frequencies for the alkali metals than for copper oxide. Thus, Millikan was able to employ radiation of frequency which would affect the alkali but not the copper oxide.

If V is the positive potential of the alkali metal necessary to prevent the escape of electrons, and f the frequency of the incident radiation, then it was found that

$$V = Kf - V_0 \quad . \quad . \quad . \quad (4\text{–}52)$$

where V_0 depends on the nature of the substance, but K has the same value for all substances. Multiplying by e and transposing, we have

$$Ve + V_0e = Kef \quad . \quad . \quad . \quad (4\text{–}53)$$

Ve is the energy with which the fastest electron escapes and, as V_0e has also the dimensions of energy, it must be interpreted as the work done by the electron in escaping from the metal. Hence, V_0e should be identical with the thermionic work function ϕe, and this is found, approximately, to be the case. K was found from the slope of the curve given by (4–52), being equal to $4{\cdot}128 \times 10^{-15}$ when V is measured in volts. The left-hand member of (4–53) is the energy with which an electron commences its flight and thus

$$\tfrac{1}{2}mv^2 = hf$$

Where h is constant, independent of the nature of the substance, and equal to Ke.

$$\text{Therefore} \qquad h = 4{\cdot}774 \times 10^{-10} \times \frac{4{\cdot}128 \times 10^{-15}}{300}$$

$$= 6{\cdot}56 \times 10^{-27} \text{ erg. sec.}$$

which is equal to the value of Planck's constant.

EINSTEIN'S THEORY: LIGHT QUANTA

The quantity hf has been termed by Einstein a "light-quantum," who put forward the theory that radiation travels in bundles of this magnitude. Referring to (4–53) we see that if f is equal to a certain value f_0 then

$$V_0e = Kef_0 \quad . \quad . \quad . \quad (4\text{–}54)$$

and below this value there will be no emission. f_0 is known as the critical or threshold frequency, and is evidently proportional to the work function of the metal. From (4–53) and (4–54)

$$Ve + Kef_0 = Kef$$

or

$$Ve = h(f - f_0)$$

which is Einstein's equation. When $\lambda = 4000$ Å, i.e. the shortest wavelength of visible light, the threshold condition is $\phi = 3$ volts. It follows that metals with higher work functions than 3 electron-volts will respond only to ultra-violet radiation. Hence, common metals, which have work functions of about 4 to 5 electron-volts, are useless as photo-electric cathodes for visible light.

According to Einstein's theory, when radiation falls on a metal the electrons receive quanta of value hf so that escape is possible, providing $hf > \phi e$. However, when an explanation of the photo-electric effect is sought in terms of the classical wave theory, grave difficulties arise. For example, a photo-electric effect can be observed when the energy falling per second per square centimetre of a substance is less than 1 erg, which is equivalent to a standard candle at a distance of 2 metres. Under these circumstances the energy of each electron liberated can be greater than 10^{-12} erg. On the wave theory an electron cannot absorb more energy than falls on the area of the molecule in which it is contained. The area of a molecule is of the order of 10^{-16} cm.^{2} and thus with 1 erg per cm.^{2} per sec. the energy received by a molecule per second is 10^{-16} erg. Thus, for an electron to acquire an energy of 10^{-12} erg would require about 10,000 sec. or 3 hours, and this time should elapse from the moment of applying the radiation until the photo-electric effect commences. Actually, however, the effect appears to occur simultaneously with the application of the radiation. This is quite possible according to Einstein's theory, for if radiation travels in quanta, and an electron can absorb a whole quantum instantaneously, there can be a photo-electric effect, resulting in the emission of a single electron as soon as the total energy of the light emitted is equal to one quantum. For the source considered, the rate of energy emission is $4\pi \times 200^2 = 5 \times 10^5$ ergs per sec. Therefore, only 2×10^{-16} sec. elapse before a quantum is emitted, which explains the almost instantaneous photo-electric effect.

PHOTO-ELECTRIC EFFECT IN NON-METALS

The photo-electric effect occurs with non-metals, but here the threshold frequency is much higher than with metals. In the case of non-conductors there are no free electrons, and hence a quantum has not only to extract an electron through the surface of the substance, but also from its parent atom. Now $hf_0 = V_0e$, or $V_0 = hf_0/e$, and thus V_0 must exceed the ionization potential of the atom by an amount equal to the work function. Observations on non-conductors are difficult, as it is not possible to conduct the charge away from the substance by means of a wire. Hence the charge collects on a non-conductor until it terminates the electronic emission.

In the case of gases the photo-electric effect takes the form of ionization, as the ejection of a photo-electron leaves the gas molecule with a positive charge. There being no surface effect, it may be anticipated that ionization should commence for $hf_0 = V_0e$ where, in this case, V_0 is the ionization potential. The maximum wavelength should also be $\lambda = hc/V_0e$. Actually, however, it is found that radiation of longer wavelengths than this may cause ionization in gases. This it appears that the effect is associated with the excitation rather than the ionization potentials of the molecules. Possibly two excited molecules may collide and their joint energy may be sufficient to produce a photo-electron from one of them.

Secondary Emission

In dealing with gaseous electrical discharges it was shown that the discharge is probably maintained by the emission of secondary electrons from the cathode. In general, electrons tend to be emitted from a metal surface whenever it is subjected to bombardment by electrons, positive ions, or metastable atoms. The motion of the latter does not constitute a current and hence knowledge of secondary emission due to this cause is slight. However, far more important are the emission processes due to the two first-mentioned causes.

SECONDARY EMISSION DUE TO ELECTRONS

When a stream of electrons strikes a metal surface, reflexion of the impinging electrons may occur, due to collisions of electrons with atoms. Cumulative reflexion may occur up to angles of 90°. In addition to this, electrons may be released from the metal, due to energy transfer from the electrons in the primary beam. It is electrons emitted in this manner which constitute secondary emission, the electrons being termed "secondary" electrons. In practice difficulty is experienced in determining the ratio of the secondary to primary electrons, as there is no precise method of distinguishing between electrons belonging to each class. It appears that the ratio is much less when a metal is freed from gas. Some figures, due to Petry, for pure metals are given in Table 4–2.*

TABLE 4–2

Metal	Maximum Ratio	Velocity of Primary Electron : Volts
Fe	1·3	348
Cu	1·32	240
Mo	1·30	356
Ni	1·3	455
W	1·45	700
Au	1·14	330

The practical importance of secondary emission will appear in dealing with valves where metals have been sought with low secondary emission. In certain apparatus, however, such as the electron multiplier, high secondary emission is desirable and may be obtained from materials usually employed as photo-electric cathodes.

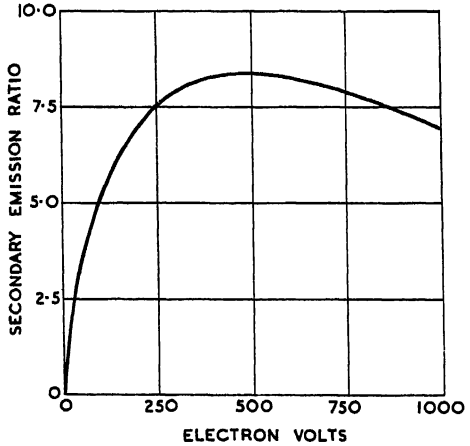


FIG. 4–17

For example, the ratio of secondary to primary electrons is greatly increased by the presence of an alkali or electro-positive metal on the surface of a pure metal. For such composite surfaces, secondary emission ratios as high as 15 have been found. This ratio for a surface which may be described as Cs–CsO–Ag (that of the electron-multiplier shown by Fig. 18–18) is shown as a function of the energy of the primary electrons by Fig. 4–17. It is possible to give some explanation of the character of this curve. When the primary electrons arrive at the surface with low energy, the energy which they can transfer to the secondaries is small, with the result that few of the latter can overcome the surface potential barrier and escape. Increasing the energy of the primaries at this stage results in more energetic secondaries and an increase in the secondary emission ratio. As the energy of the incident electrons is continually increased, however, they penetrate deeper into the metal and the secondaries produced must travel further in order to reach the surface. In doing so, the number of collisions made by the secondaries increases, with the result that their random velocity is increased at the expense of their x-directed velocity. Thus, as the energy of the primaries is increased, the secondary emission ratio must pass through a maximum as shown.

---

* *Phys. Rev.* 26 (1925), p. 346; 28 (1926), p. 362.

The energy of the majority of the secondary electrons is usually not more than about 3 electron-volts. This is, perhaps, to be expected, for a high-energy secondary within the metal should be able to act in a similar manner on other electrons as a primary, and thus gradually lose its energy.

SECONDARY EMISSION DUE TO POSITIVE IONS

In practice, secondary emission due to positive ion bombardment is usually far more important than that due to primary electrons. This is because the field at the cathode, where the positive ions impinge, is in such a direction as to remove the secondary electrons, whereas the field at the anode restricts the escape of electrons. As we have already seen, according to one theory, secondary emission at the cathode is essential to a self-maintained discharge.

A positive ion on arrival at the cathode possesses two forms of energy: potential energy equal in amount to its ionizing potential, and kinetic energy due to its velocity. On neutralization of its charge the energy of an ion may be absorbed in various ways. Thus, energy is necessary to extract the neutralizing electron from the cathode and if, in addition, secondary emission occurs, energy will be needed to extract the secondary electrons. Heat is also produced. Suppose the total current at the cathode is 1 amp. and of this x amp. is due to positive ions and $(1 - x)$ amp. to electrons. Let V be the potential difference between the cathode and its surroundings, V_i the ionizing potential of the gas, and ϕ the work function voltage of the cathode. If the kinetic energy of the ions corresponds to a fall through a potential difference V', then, assuming the secondary electrons have negligible initial velocities,

Potential energy of ions $= xV_i$
Kinetic energy of ions $= xV'$
Energy for extraction of neutralizing electrons $= x\phi$
Energy for extraction of secondary electrons $= (1 - x)\phi$

Therefore balance of energy possessed by ions is

$$\begin{aligned} E &= x(V' + V_i - \phi) - (1 - x)\phi \\ &= x(V' + V_i) - \phi \qquad . \quad . \quad . \quad . \quad (4\text{–}55) \end{aligned}$$

If, after neutralizing, the ion is devoid of energy, then (4–55) is the rate of heat production at the cathode in watts per ampere. Hence under suitable circumstances x may be determined. In some experiments, due to Oliphant,* conditions were arranged so that V' was equal to V and ϕ and V_i were relatively small. In this case (4–55) becomes

$$E = xV \qquad . \quad . \quad . \quad . \quad (4\text{–}56)$$

x was determined from (4–56) by measuring the additional heat needed to maintain the cathode at a constant temperature as V was reduced. It was found that x was approximately constant for values of V between 40 and 600 volts. Some values of x for various gases are given by Table 4–3.

TABLE 4–3

GAS	x
He	0·7
A	0·8
Ne	0·55
H_2	0·7
Hg	0·75

Conduction in Metals, Insulators, and Semi-conductors

In view of the Fermi-Dirac theory of the metallic state it is necessary to reconsider the subject of electrical conductivity. It appears that electrons with energies well below the normal maximum level cannot take part in electrical conduction. When an electron is under the influence of an electric field it tends to change its energy state. As the states adjacent to those of electrons having energy less than E_m are full, it is apparent that such electrons cannot take part in electrical conduction. However, those relatively few electrons with energy equal to E_m or above may be forced up into an unoccupied state or level, such states being known as *running* states or levels. This term is employed because it is probable that the flow of electrons within a conductor takes place principally by means of the motion of electrons in these levels.

---

* *Proc. Roy. Soc.* A, 132 (1931), p. 631.

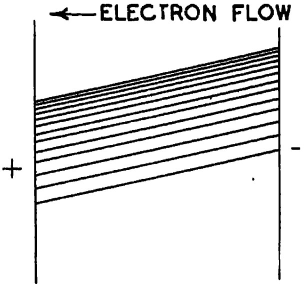


FIG. 4–18

If a potential difference is applied to two opposite faces of a conductor, this has the effect of tilting all the energy levels into an inclined position. Thus, considering Fig. 4–18, if the right-hand side of the conductor is made negative relatively to the left, the energy of the electrons to the right will be increased. This increase in energy is initially potential, and the electrons can now move to the left, passing into higher kinetic-energy states at the expense of potential energy. Obviously the electrons below E_m cannot move because the levels immediately above them are full. The energy of a conducting electron does not remain constant because it collides with the lattice structure, thus losing energy in the form of heat. Such losses constitute the resistance of the conductor.

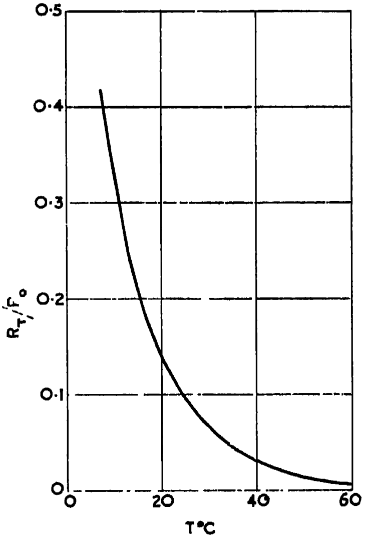


FIG. 4–19

INSULATORS AND SEMI-CONDUCTORS

Within an insulator the electrons reside in definite kinetic-energy levels exactly as within a conductor. However, unlike a conductor, the electrons within an insulator are firmly attached to their parent atoms and hence cannot readily accept additional energy to force them into unoccupied levels. Thus, an applied potential may steeply tilt the energy levels without electrons being able to pass from one level to another, this resulting in an absence of conduction.

The terms "conductor" and "insulator" are, of course, only relative, for it is doubtful whether a perfect conductor or insulator exists. Most insulators have some slight conductivity which tends rapidly to increase with an increase in temperature. Reference to Fig. 4–19 shows this property for red fibre in a dry state. The effect of heat on an insulator is to increase the kinetic energy of its electrons and so force some of them into running levels, thus leading to an increase in conductivity.

A semi-conductor may be regarded as a composite of a conductor and an insulator. The conditions may be illustrated by Fig. 4–20, in which it will be seen that running levels exist above a thin layer of filled levels, the latter in turn residing above a no-level band. The thin layer is said to consist of "impurity" levels, as it is believed to contain electrons belonging to impurities within the substance. Typical important examples of semi-conductors are cuprous oxide and selenium. In the former it is believed that the impurity levels are associated with an excess of oxygen, i.e. oxygen beyond that demanded by the chemical formula of cuprous oxide. Certainly the semi-conducting properties are due to oxygen excess, for pure cuprous oxide has a specific resistance of the order of 10^8 ohm per cm.^{3}, while that with an excess of oxygen has a specific resistance of the order of 10^3 ohm per cm.^{3}

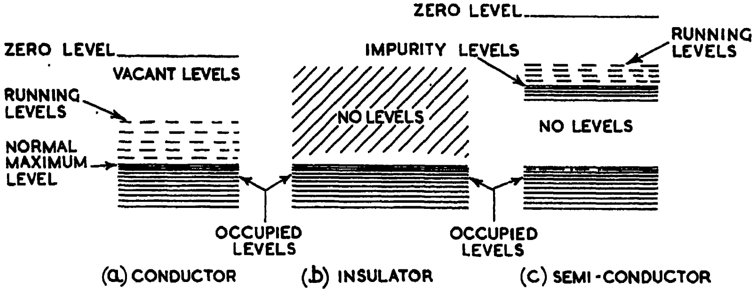


FIG. 4–20

The conductivity of a semi-conductor is limited by the comparatively few electrons in the running levels. This is probably due to the lack of depth of the impurity levels which supply the running levels. The conductivity of a semi-conductor is extremely susceptible to temperature variations, the explanation of this being the same as for an insulator, i.e. an increase in temperature forces electrons from both the impurity levels and those below the no-level band up into running levels, the first of these two effects having much the greater result on the increase in conductivity.

CHAPTER V

X-RAYS

It has already been stated in Chapter III that if a gas-discharge tube is exhausted to the degree that the Crookes dark space fills the tube, a radiation, termed *X*-radiation, is emitted from wherever cathode rays (or electrons) strike. The discovery of this form of radiation was due to Röntgen, and the first indication of its existence may be said to be evidenced by the fluorescence of the glass tube when this is sufficiently exhausted. By covering the tube with black paper, Röntgen found that fluorescence could still be excited in a barium platino-cyanide screen placed at some distance from the tube. From this he concluded that some hitherto unknown form of radiation existed, capable of penetrating opaque substances and producing effects at distances remote from its point of generation. The capability of X-rays to penetrate solid bodies is, of course, a fact of the greatest importance, and this is further enhanced by the fact that they are capable of affecting a photographic plate. From these two phenomena arise the science of radiography.

Methods of Production

The original tube with which Röntgen made his discovery of X-rays was not such as is used for their production to-day. This was in fact what is usually termed a Crookes tube, the rays from such a tube being "soft" in character. X-rays are classified into "soft" and "hard" according to their capacity for penetrating substances, the hard rays being relatively more penetrating than the soft. In order to produce harder rays than those obtainable from a Crookes tube, early X-ray tubes took the form shown by Fig. 5–1. This possessed a concave cathode, the purpose of which is to focus the electrons on to a metal target *T*. The rays arise from *T* in the manner shown. That the rays originate from where the electrons (or cathode rays) strike may be demonstrated in the manner shown by Fig. 5–2, where *B* is a photographic plate and *A* a metal plate possessing a number of holes. If an exposure is made, *B* developed and replaced in its initial position, then, producing the lines backwards from the images through the holes, it will be found that they converge on a common point on the target, thus showing that the X-rays arose from this point.

SECONDARY RADIATION

When X-rays fall on any material, further rays are produced, termed *secondary radiation*. This radiation is complex in character, consisting mainly of secondary X-rays. These rays may be classified under two headings, namely scattered X-rays and characteristic X-rays. The former consist of rays similar to those of the primary beam, and may be considered as rays which have been deflected by the substance upon which they have fallen. The characteristic X-rays are so termed because they are characteristic of the substance on which the primary rays have fallen.

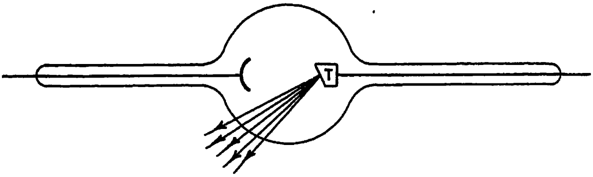


FIG. 5–1

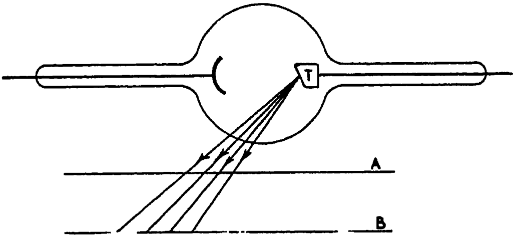


FIG. 5–2

Origin of X-rays

CLASSICAL THEORY

The first explanation of the origin of X-rays was due to Stokes, who suggested that the sudden stoppage of cathode rays (i.e. electrons) when they strike a solid body causes the emission of electromagnetic pulses or radiation of very short wavelength. However, earlier than this J. J. Thomson had observed that if cathode rays were rapidly moving charged particles (as he afterwards proved to be the case) then, in accordance with Maxwell's theory of radiation, they should produce radiation if suddenly stopped.

In order to pursue this idea we shall consider an electron moving along AB, Fig. 5–3, with a velocity v, which is small compared with that of light. The magnetic field due to the electron is uniformly distributed around it and, of course, accompanies the electron while, and only while, it is in motion. It will now be assumed that when the electron reaches O a force acts upon it which brings it to rest in a short time dt, it coming to rest at O'. Considering the field distribution at a time t after the electron is at rest, if a sphere of radius $r = ct$ is described about O', a disturbance will have passed over the field within this sphere, the field originating from (or terminating on) O'. Now, the field distribution outside a sphere of radius $c(t + dt)$, described about O, will be the same as if the electron had not been arrested, because the disturbance can only travel outwards from O with velocity c. If the electron continued to move uniformly, the field would have been uniformly distributed about O'' where $OO'' = v(t + dt)$. The disturbance due to the stoppage of the electron will accordingly be contained between the two spheres. The form of the field in this region cannot be predicted, but it is evident that it will possess a tangential component, this travelling outwards with the velocity of light and constituting the X-rays. If vdt is small compared with cdt, the spheres may be regarded as concentric and the disturbance will be confined between two spheres of radii ct and $c(t + dt)$, the thickness of the disturbance being cdt. Resolving the disturbance into two right-angled components, we have

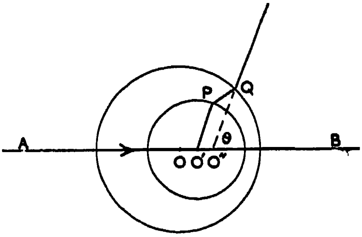


FIG. 5–3

$$\frac{\text{Tangential Electric Displacement}}{\text{Normal Electric Displacement}} = \frac{vt \sin\theta}{cdt}$$

The normal displacement is $e/4\pi r^2$ and, as $r = ct$, the tangential component is

$$\frac{e}{4\pi c^2 t^2} \cdot \frac{vt \sin\theta}{cdt}$$

or

$$\frac{ev \sin\theta}{4\pi r c^2 dt} \quad . \quad . \quad . \quad . \quad . \quad (5\text{–}1)$$

where e is expressed in e.s.u.

As the latter is moving at right-angles to its direction with velocity c, it will produce a magnetic field H equal to $4\pi c/c$ times the tangential displacement,* i.e.

$$H = \frac{ev \sin \theta}{rc^2 dt} \quad . \quad . \quad . \quad . \qquad (5\text{–}2)$$

From (5–1) and (5–2) it will be noted that the electric and magnetic field strengths vary inversely as the distance from the electron, whereas the normal component of the field varies inversely as the square of that distance. Hence the intensity in the disturbance, except in the vicinity of the electron, will be large compared with the intensity outside it. Thus, a pulse of electromagnetic disturbance travels outward from the electron, behaving in some respects like ordinary light. The principal differences are, however, that the thickness of the pulse is small compared with the wavelength of light and lacks that regular periodic character distinguishing a train of waves of constant wavelength.

ENERGY OF RADIATION

According to Maxwell's theory, the energy in an electromagnetic wave is equally divided between the magnetic and electric fields. Hence the energy in the pulse is given by either

$$2 \times \frac{\mu H^2}{8\pi} \text{ or } 2 \times \frac{kX^2}{8\pi} \text{ ergs per unit volume,}$$

where X is the electric field strength and μ and k the permeability and specific inductive capacity of the medium in which the radiation occurs. Assuming $\mu = 1$, the energy in the pulse per unit volume is

$$\frac{1}{4\pi}\left(\frac{ev \sin \theta}{rc^2 dt}\right)^2$$

The volume of shell included in radii making angles θ and $\theta + d\theta$ with the axis AB is $cdt \,.\, 2\pi r \sin \theta \,.\, rd\theta$, and the total energy is therefore

$$2\int_0^{\pi/2} 2\pi r^2 \sin \theta c dt \,.\, \frac{1}{4\pi}\left(\frac{ev \sin \theta}{rc^2 dt}\right)^2 d\theta$$

$$= \frac{2}{3} \cdot \frac{e^2 v^2}{c^3 dt}$$

---

* *Electricity and Magnetism*, p. 421, S. G. Starling (Longmans, Green).

But this is the energy radiated in time dt, so that the rate of radiation is

$$\frac{2}{3} \cdot \frac{e^2v^2}{c^3(dt)^2} \text{ ergs per sec.}$$

It may be noted that $(v/dt)^2$ is the average rate of acceleration of the electron. Writing this as f^2, the rate of radiation is

$$\frac{2}{3} \cdot \frac{e^2f^2}{c^3} \text{ ergs per sec.}$$

and this is true whether the electron is brought to rest or merely undergoes a change of velocity.

QUANTUM THEORY

Although the foregoing theory of the origin of X-rays is not without merit (as we shall see later when dealing with scattering), it is more probable that explanation of X-rays is to be found in the quantum theory. According to this theory radiation could be produced by a bombarding electron penetrating an atom with a consequent displacement of a planetary electron to an orbit more remote from the nucleus. Under such circumstances the atom would be in an excited state and, on return of the displaced electron to the ground state, radiation would occur.

ABSORPTION OF X-RAYS

Reference has been made previously to the variation in the penetrating power of X-rays and their classification into hard and soft rays. The degree of hardness is largely associated with the potential difference through which an electron falls in passing from the cathode to the target being greater the larger the potential difference. The penetrating power of the rays is obviously of such importance that it is necessary to have some means of measuring this quality. In practice this is effected by measuring the degree to which the rays are absorbed by a suitable screen.

Now, if a beam of X-rays of intensity I pass through a length dx of any material, the loss in intensity over dx may be written dI. For a given material the amount of the absorbed intensity may be written $\mu I dx$, where μ is termed the absorption coefficient, or the fraction of the intensity of the beam absorbed per unit length of path. Thus

$$-dI = \mu I dx$$

$$\frac{dI}{I} = -\mu dx$$

and

$$\log I = -\mu x + K$$

If I_0 is the intensity of the incident beam, then when $x = 0$, $I = I_0$.

Hence $$K = \log I_0$$

and $$\log I = -\mu x + \log I_0$$

from which $$I = I_0 \varepsilon^{-\mu x} \quad . \quad . \quad . \quad . \quad (5\text{–}3)$$

In order to measure the absorption coefficient the apparatus shown by Fig. 5–4 may be employed. This consists of an ionization chamber before which is placed a sheet of material, S, in which absorption is to be produced. The ionization chamber consists of an insulated aluminium plate mounted within an aluminium box, this metal being relatively easily penetrated by X-rays. A potential difference is maintained between the box and plate, sufficient to produce a saturation current in the gas when ionization occurs. The magnitude of this current is a measure of the ionizing power and thus of the intensity of the beam. By measuring the saturation current with and without S in position, I/I_0, and thus μ, may be found.

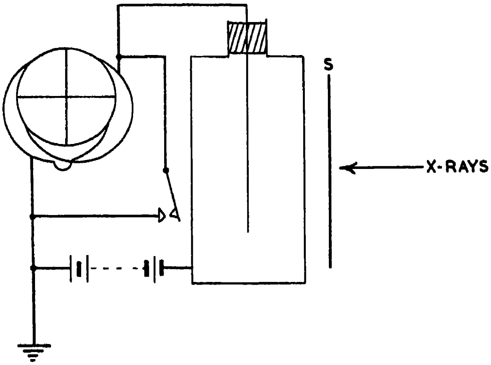


FIG. 5–4

Provided the X-rays being absorbed by a given material are homogeneous, the law expressed by (5–3) is found to be true. Where the X-rays are heterogeneous, the law is not obeyed, and plots of $\log I/I_0$ against x do not give straight lines. However, the law may be approximated to by adopting a series of straight lines, up to four in number.

As X-rays are classified by their absorbability in different substances, it was found that a better measure of the quality of the rays than the absorption coefficient is what is termed the mass-absorption coefficient, i.e. the ratio of the absorption coefficient to the density of the substance, viz. μ/ρ. For a sheet of absorbing material of area A, volume V, mass m, and density ρ, we have

$$V = Ad$$

$$m = V\rho$$

from which

$$d = m/A\rho$$

Substituting for x in (5 3).

$$I = I_0 e^{-\frac{\mu}{\rho}\left(\frac{m}{A}\right)}$$

where μ/ρ is the mass-absorption coefficient. Thus, if an absorber is described in terms of m/A, i.e. grams per cm.^{2}, and $\mu \propto \rho$, the mass absorption coefficient will be independent of the density of the substance.

In practice it is found that μ/ρ is not constant, for μ increases with the atomic weight of the substance more rapidly than proportionally to the density. In addition, μ is a function of the "hardness" of the rays (the hardness increases with the voltage at which the rays are generated), decreasing as the hardness increases. Table 5 1 indicates the manner in which μ decreases as the X-ray tube potential increases.

TABLE 5 1

VOLTS	μ
10,000	0·0018
30,000	0·001
60,000	0·0004
90,000	0·00029

Scattering: Classical and Quantum Theories

As stated previously, when X-rays fall on a material scattering occurs. Although the scattered radiation consists of X-rays belonging to the primary beam, these rays after scattering are somewhat modified, usually softened. The modification results from an absorption and re-emission of the rays, fluorescent radiation* and what is known as *Compton scattering* being produced. Thus, the measurement of absorption includes true absorption (i.e. conversion of X ray energy into heat) and scattering. Different substances do not produce the same degree of scattering, those of high atomic weight producing a greater effect than those of low atomic weight.

---

* See Chapter VII.

If an electromagnetic disturbance due to an accelerated electron falls on another electron, the latter will be subjected to an electric field X and will thus receive an acceleration

$$f_1 = \frac{Xe}{m} = \frac{e^2 f \sin\theta}{mrc^2}$$

The electron will, therefore, give out radiation, which (since f_1 is proportional to f) will have the same quality as the radiation exciting it. If there are n atoms per cm.^{3} of a substance on which an X-ray beam falls, and there are Z loosely bound electrons per atom, then the possibility exists of setting in vibration the Z electrons. The latter will then scatter by re-radiation a fraction of the original energy in all directions, with maximum intensity in a plane perpendicular to the direction of electron oscillation. As the rate of radiation of an accelerated electron is $\frac{2}{3}e^2f^2/c^3$, then 1 cm.^{3} of a substance on which an X-ray beam falls will radiate energy at a rate

$$S = \frac{2}{3}\frac{e^2f^2}{c^3}nZ \text{ ergs per sec.}$$

this being in the form of scattered radiation. The rate at which energy is being received per square centimetre from the incident X-ray beam is

$$P = \frac{cX^2}{4\pi} \text{ ergs per sec.}$$

But $X = mf/e$, and hence

$$P = \frac{cm^2f^2}{4\pi e^2}$$

Thus

$$\frac{S}{P} = \frac{8\pi}{3}\frac{e^4}{m^2c^4}nZ$$

Now $S/P = \mu$, the absorption coefficient. If instead of μ the mass absorption coefficient is employed, then

$$\frac{\mu}{\rho} = \frac{S}{\rho P} = \frac{8\pi}{3}\frac{e^4}{\rho m^2c^4}nZ$$

But $\rho = nm_HA$, where m_H is the mass of a hydrogen atom and A the atomic weight of the scatterer. Hence

$$\frac{\mu}{\rho} = \frac{8\pi}{3}\frac{e^4}{nm_HAm^2c^4}nZ = \frac{8\pi}{3}\frac{e^4}{m^2c^4}\frac{1}{m_H}\frac{Z}{A} = K\frac{Z}{A}$$

where K has a value of about 0·4. Thus

$$\frac{\mu}{\rho} = 0{\cdot}4\,\frac{Z}{A}$$

Experimental measurements of μ/ρ for the lighter elements from H to Al give 0·2. Hence

$$Z = A/2$$

which indicates that the number of electrons within an atom is about one-half of its atomic weight. Reference to a table of atomic weights shows that this is approximately true.*

QUANTUM THEORY OF SCATTERING

According to the classical theory of scattering given above, the frequency of a scattering electron should be the same as that of the electric field under whose influence it oscillates. However, as previously stated, the scattered rays are usually softened, which means that they are of longer wavelength than those of the primary beam. The change of wavelength was investigated by A. H. Compton, who found that it could be explained by an application of the quantum theory.

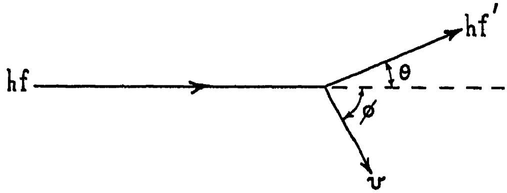


FIG. 5–5

Compton considered X-rays as light quanta or photons which collided elastically with electrons within the atom. The photons were considered to have momentum, thus setting in motion the electrons upon collision. According to (3–24) and (3–25) the momentum and mass of the incident photon are respectively hf/c and hf/c^2. As a result of a collision between a photon and an electron the latter emits a quantum hf' of scattered rays, making an angle θ with the direction of the incident rays, as shown by Fig. 5–5. Also, the electron, known as the recoil electron, is set in motion with an initial velocity v in a direction making an angle ϕ with that of the incident rays. Now, according to (3–28), the kinetic energy of the electron is

---

* *Physical and Chemical Constants*, Kaye and Laby (Longmans, Green).

$$m_0c^2\left[\frac{1}{\sqrt{1-v^2/c^2}}-1\right]$$

and from the law of conservation of energy we have

$$hf = hf' + m_0c^2\left[\frac{1}{\sqrt{1-v^2/c^2}}-1\right] \quad . \quad . \quad (5\text{–}4)$$

Also from the law of conservation of momentum

$$\frac{hf}{c} = \frac{hf'}{c}\cos\theta + \frac{m_0v\cos\phi}{\sqrt{1-v^2/c^2}} \quad . \quad . \quad . \quad (5\text{–}5)$$

$$0 = \frac{hf'}{c}\sin\theta - \frac{m_0v\sin\phi}{\sqrt{1-v^2/c^2}} \quad . \quad . \quad . \quad (5\text{–}6)$$

these two equations, respectively, referring to the x components along the beam and the y components normal to the beam. The mass of the electron is given by (3–27). From (5–4), (5–5), and (5–6), we have

$$\lambda' = \frac{c}{f'} = \lambda + \frac{h}{m_0c}(1-\cos\theta)$$

or

$$\lambda' - \lambda = \frac{h}{m_0c}(1-\cos\theta) = \frac{2h}{m_0c}\sin^2\frac{\theta}{2}$$

Also

$$\tan\phi = \frac{\cot\frac{\theta}{2}}{1+\frac{hf}{m_0c^2}}$$

These results indicate that λ' is greater than λ unless $\theta = 0$. The difference $(\lambda' - \lambda)$ is evidently a function of the scattering angle θ and is independent of λ. Thus, the percentage change in wavelength is large for small wavelengths, but small for large ones. Substituting for h, m_0, and c, we have

$$\lambda' - \lambda = 0{\cdot}0242(1-\cos\theta)$$

The foregoing results are confirmed by experiment and appear to indicate that in some respects X-rays behave as particles of energy hf and momentum hf/c. However, in other respects X-rays behave as radiation of extremely small wavelength, this apparent inconsistency being as yet unresolved. A similar anomaly has, of course, already been found in Chapter III in dealing with photoelectricity.

Secondary Corpuscular Radiation

In addition to recoil electrons, a beam of X-rays incident on a substance also liberates photo-electrons. The latter are not emitted equally in all directions, but in the case of a thin scatterer, such as gold leaf, considerably more electrons (or corpuscular radiation) are emitted from the emergence side than from the incident side. Thus, the electrons are projected from the atoms with a component velocity in the direction of the primary beam. The electric field and force in the primary beam are at right-angles to the direction of propagation and hence the electrons tend to be ejected in a direction perpendicular to that of the beam. However, because of the magnetic field in the beam the electrons after ejection will be acted on by this and the resultant motion will then be in the direction of the beam.

If v is the velocity of electron projection and u the component velocity along the beam, then, if the electron absorbs a photon hf,

$$mu = \frac{hf}{c}$$

$$\tfrac{1}{2}mv^2 = hf$$

and

$$\frac{u}{v} = \frac{v}{2c}$$

For photo-electrons projected by X-rays, v is of the order of 10^{10} cm. per sec. Hence $u/v = 1/6$, which is approximately the value found from experiment.

Characteristic X-rays

Of the two kinds of secondary X-rays produced when a primary beam falls on a scattering substance, one of these, the characteristic or fluorescent radiation, depends on the nature of the substance and not on the quality of the primary beam. However, to produce any characteristic X-ray, the primary X-rays must be at least as hard as the characteristic X-rays produced. This means that the atomic weight of the anticathode in the X-ray tube must exceed that of the scatterer and also that the potential of the tube must exceed a certain value. These facts indicate that the characteristic X-rays are due to some electronic conditions within the atom.

Employing absorbing screens of aluminium, Barkla measured the mass absorption coefficient for the characteristic X-rays arising from a variety of materials when placed in a beam of X-rays. The characteristic rays tended to fall into two distinct groups, as shown by Fig. 5–6, where log μ/ρ is plotted against the logarithm of the atomic weight of the scattering element. These two groups were termed the series K and series L fluorescent radiations, the former being more penetrating and less absorbed than the latter.

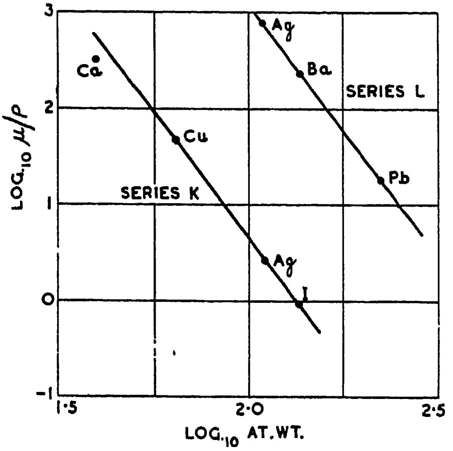


FIG. 5–6

If an absorbing material is irradiated by a beam of homogeneous radiation insufficiently hard to excite the characteristic L radiation of the absorbing material, it is found that the absorption coefficient diminishes rapidly with diminishing wavelength* of the homogeneous radiation, the relation being expressible by $\mu/\rho = K\lambda^3$. This is shown by the branch M of the curve of Fig. 5–7. As the wavelength is continually decreased, a value is ultimately reached at which the absorption suddenly rises to a relatively high value. At this wavelength the L radiation of the absorbing material is excited and the energy of this characteristic radiation is abstracted from that of the primary beam. Beyond L, Fig. 5–7, the absorption again decreases in accordance with a third power law until the wavelength is such that the K radiation is excited. For elements of high atomic weight an M discontinuity exists as well as those of the L and K series. The absorption curve for varying wavelengths may be written

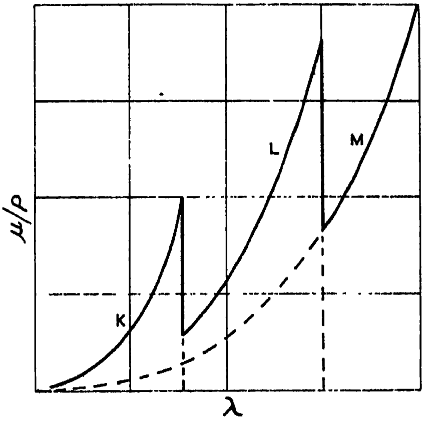


FIG. 5–7

$$\frac{\mu}{\rho} = (K_K + K_L + K_M + \ldots)\lambda^3$$

---

* See p. 210.

where K_K, K_L, K_M, etc., are constants for a given absorbing material, each of which becomes effective when the incident wave-length is less than that corresponding to the K, L, M, etc., absorption limits.

A similar curve to that of Fig. 5–7 occurs if μ/ρ is plotted against atomic weight for a given quality of radiation. As the atomic weight increases the absorption also increases, while the atomic weight is such that the *K*, *L*, and *M* radiations can be absorbed. Beyond the element of the same material as that responsible for the radiation, radiation can no longer be absorbed by the *K* structure of the elements. Hence from this point the *K* absorption ceases, and only the *L* and *M* absorptions for the elements of increasing atomic weight occur. As the atomic weight increases, the absorption coefficient again increases until the hardness of the *L* radiation of a certain element exceeds that of the incident radiation. At this point the *L* absorption will abruptly cease, as shown by Fig. 5–8.

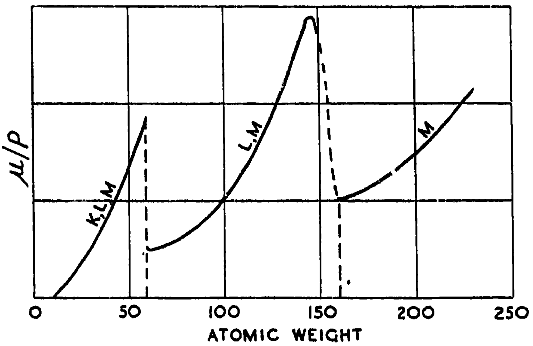


Fig. 5–8

The foregoing phenomena indicate that each element possesses a series of X-ray characteristics designated by the letters *K*, *L*, *M*, *N*, etc. These characteristics differ in the hardness of the rays required to excite them, the hardness diminishing with the advancing position of the letter in the alphabet. The series can be invoked for all elements, but the higher the atomic weight the greater must be the hardness of the primary exciting beam. Thus a beam which may be adequate to produce the *K* series in a light element may only be able to produce the *M* series in a heavier element. In fact, the *L* and *M* series in some of the lighter elements are not X-rays, but far ultra-violet radiation. Contrasting Al and W, an X-ray tube with electrons having an energy of $1500eV$ may excite the K series for Al, but $60{,}000eV$ are required to excite the same series for W.

It is evident that characteristic X-rays are closely connected with atomic weight or number, and hence with atomic structure. The K series first appears in elements of low atomic weight. As the atomic weight increases, the L series next appears, and later the M series, and so on. On page 201 it was briefly suggested that the most probable explanation of the origin of X-rays was to be found in the quantum theory rather than in the classical theory. This suggestion is adequately supported by the fact that when means were found for measuring the frequencies of X-ray spectra, the latter were found to be connected with the accelerating potential of the X-ray tube by the relation

$$Ve = hf$$

where V is the accelerating potential and f the frequency of the corresponding series produced.

Let it be assumed that for the most firmly bound electrons within the atom (i.e. those with the principal quantum number $n = 1$) E_K is the work necessary to displace one of these from its normal orbit. The lowest frequency which incident radiation may have to effect this displacement is $f_K = E_K/h$. When an increasing frequency reaches this critical value, absorption by the K electrons commences with a marked increase in the absorption coefficient. Hence the K absorption discontinuity indicates the point at which the K shell electrons are capable of abstracting energy from the incident beam. From the atomic theory given in Chapter I it should now be clear how the various X-ray series arise. Thus, the K series corresponds to $n = 1$, the L series to $n = 2$, and the M series to $n = 3$. Also the comparatively few series indicate that only a few energy levels or electron orbits exist within the atom.

Let it be supposed that an electron has been ejected from the K shell. The vacancy created will be filled by another electron, generally coming from one of the outer shells. If this electron originates from the L_1 level, energy will be liberated equal to

$$hf = E_K - E_{L1} = hf_K - hf_{L1}$$

Thus, f should be the frequency of a line in the K series of the element. For tungsten the values of the K and L discontinuities, respectively, occur at wavelengths of 0·1785 Å and 1·2136 Å.

Writing $f = c/\lambda$

$$\frac{hc}{\lambda} = \frac{hc}{\lambda_K} - \frac{hc}{\lambda_{L1}}$$

or

$$\lambda = \frac{\lambda_K \lambda_{L1}}{\lambda_{L1} - \lambda_K}$$

$$= 0{\cdot}2086 \text{ Å}$$

Measurement shows the $K_{\alpha 1}$ line to have this value.

It will have been noted above that subscripts are attached to K and L. This is because the X-ray spectral lines of any element can be arranged in series or groups of lines, each line having a slightly different wavelength from the others. The series having the shortest wavelengths is the K series, that with the next shortest, the L series, and so on. The K series usually consists of four lines known as $K_{\alpha 1}$, $K_{\alpha 2}$, K_β, and K_γ. Table 5–2* gives the wavelengths of the lines of the K series for several elements. It will be noted that the higher the atomic number the shorter the wavelengths become. The various lines are associated with the different electron

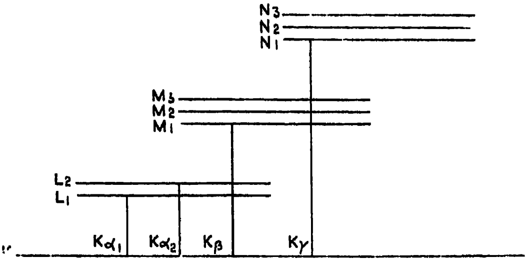


FIG. 5–9

TABLE 5–2

ELEMENT	ATOMIC NUMBER	$K_{\alpha 2}$	$K_{\alpha 1}$	K_β	K_γ
Na	11	—	11·883	11·591	—
K	19	3·738	3·733	3·446	-
Fe	26	1·932	1·9324	1·754	1·736
Br	35	1·040	1·035	0·929	0·914
Rh	45	0·616	0·612	0·545	0·534
Cs	55	0·402	0·398	0·352	
W	74	0·2134	0·2086	0·1842	0·179

---

* For fuller data see *Physical and Chemical Constants*, Kaye and Laby, p. 99, (Longmans, Green).

orbits within the atom and may be represented by the energy level diagram of Fig. 5–9. It is found that there is only one K level, but there may be three L levels, and seven N levels. From Fig. 5–9 it appears that the $K_{\alpha1}$ and $K_{\alpha2}$ lines occur when an electron falls from the L levels to the K level. Also the K_β and K_γ lines occur when electrons, respectively, fall from M and N levels. For the L level, $n = 2$ and, as shown on page 48, for this number there are two possible orbits, one circular and one elliptical. It may be presumed, therefore, that the $K_{\alpha1}$ and $K_{\alpha2}$ lines originate from electrons which fall from these two orbits into the K orbit, for which $n = 1$.

X-ray Diffraction

It has been stated that X-rays consist of electromagnetic radiation of extremely short wavelengths and the method of measuring such wavelengths must now be considered. The discovery that X-rays may be diffracted by crystalline substances is the origin of the method employed and hence some attention must be given to crystal structure. Crystals are homogeneous solid substances which are bounded by plane surfaces termed the faces of the crystal. Certain angles exist between crystal faces, these angles being of constant value and independent of the size of the crystal. A crystal grows by the deposition of uniform layers upon its faces, and hence it may become larger without the angles between the faces changing. The regular structure of a crystal is due to regularly spaced geometrical arrangements of atoms or ions which form unit cells from which larger crystal aggregates are developed. The atoms or ions lie in regularly spaced planes, the distance between such planes being of the order of 10^{-8} cm. As we have already seen, when X-rays fall on a substance the latter transmits part of the rays and scatters or re-radiates the remainder. According to the classical theory of scattering, re-radiation is due to electron acceleration within individual atoms of the scattering substance. If scattering occurs from a regular geometrical arrangement of atoms, it appears possible that at some position beyond such an arrangement the scattered radiation should exhibit some orderly pattern. This is so, the waves of radiation uniting crest to crest at some places and crest to trough at others. Hence in the first case a point of maximum intensity occurs, and in the second case a point of zero intensity. By directing a fine beam of X-rays into a crystal and placing a photographic plate at the back of the latter, the plate, after suitable exposure and development, shows a regularly spaced number of spots arranged round a large central spot. The latter is, of course, due to the effect of the directly transmitted radiation. The image received by the photographic plate is termed a *diffraction pattern.*

CRYSTAL STRUCTURE

Crystals may have a large variety of shapes, but for the purpose of illustrating their application to X-rays it is convenient to consider one of the most simple and important crystals, namely rock salt. This substance crystallizes on what is known as the cubic system, a typical crystal being represented by Fig. 5–10. Referring to this, it will be noted that the crystal is made up of a number of unit cubes at the corners of which are Na^+ and Cl^- ions placed alternately. The ions may be regarded as lying in planes, and one set of planes may be taken as those lying parallel to the top plane. Thus the plane 1234 is one of this set. Another set of parallel planes may be taken through the diagonals of the elementary cubes, as shown by *AA′B′B*.

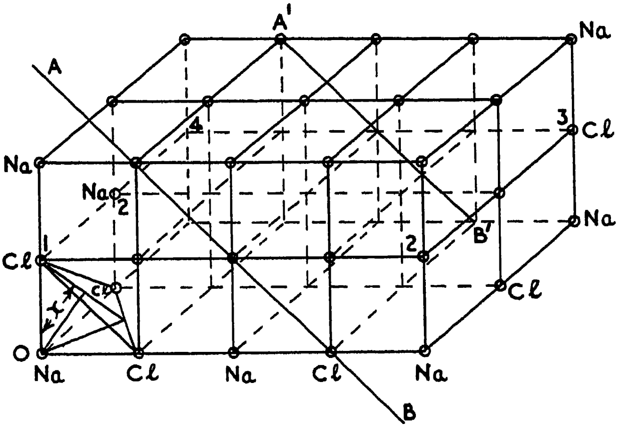


FIG. 5–10

The various sets of planes are designated by numbers known as Miller indices, the different faces of the crystal being specified by the intercepts cut off from three axes by the faces. As axes, lines are taken which are parallel to the three edges of the crystal, and which do not lie in one plane. Except for cubic crystals, the angles between the axes, x, y, and z, are different, as are also the unit lengths a, b, and c along the axes. Regarding a given plane within a crystal, this plane will have intercepts on each of the axes, x, y, z, at, say,

e, f, g. The lengths of these intercepts from the origin in terms of a, b, and c are given by numbers j, k, and l. Thus, $e = ja$, $f = kb$, and $g = lc$. The values of the numerators of the reciprocals of the numbers, j, k, l, when expressed with their least common denominator, are the Miller indices. Thus

$$\frac{1}{j} = \frac{kl}{jkl}$$

$$\frac{1}{k} = \frac{jl}{jkl}$$

$$\frac{1}{l} = \frac{jk}{jkl}$$

and the right-hand members give the Miller indices of the (kl) (jl) (jk) plane. For example, if $e = 2a$, $f = 1b$, and $g = 2c$, then $j = 2$, $k = 1$, and $l = 2$. Hence

$$\frac{1}{j} = \frac{1}{2}$$

$$\frac{1}{k} = \frac{1}{1}$$

$$\frac{1}{l} = \frac{1}{2}$$

The L.C.M. being 4, we have

$$\frac{1}{j} = \frac{2}{4}$$

$$\frac{1}{k} = \frac{4}{4}$$

$$\frac{1}{l} = \frac{2}{4}$$

and the plane is designated as the 242 plane.

In the case of cubic crystals, the axes are the rectangular coordinates, x, y, z, the unit lengths a, b, c being equal. For a plane at $e = a$ and parallel to the yz plane, $f = \infty$, $g = \infty$. Hence, $j = 1$, $k = \infty$, $l = \infty$, and the Miller indices are 1, 0, 0. Hence this plane is designated as the 100 plane. For a diagonal plane, similar to that shown by Fig. 5–10, with $e = a$, $f = b = a$, and $g = \infty$, the Miller indices are 1, 1, 0. Hence this plane is designated as the 110 plane. Again, the plane shown by 567 has intercepts

$e = f = g = a$. The Miller indices are therefore 1, 1, 1, and this plane is designated as the 111 plane.

GRATING SPACE

The distance between equally spaced parallel planes of atoms within a crystal is usually denoted by d, which is termed the crystal grating space. The spacings between the various sets of planes are simply related. Thus, referring to Fig. 5–10, it is evident that if d is the distance between two adjacent 100 planes, the distance between two adjacent 110 planes is $d/\sqrt{2}$. For the 111 planes the angles at O are reproduced by Fig. 5–11. From this,

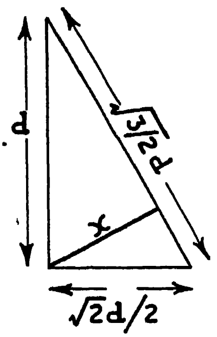


FIG. 5–11

$$\sqrt{d^2 + \left(\frac{d}{\sqrt{2}}\right)^2} = \sqrt{\frac{3}{2}}\,d$$

$$\frac{x}{d} = \frac{\sqrt{2}d/2}{\sqrt{\frac{3}{2}}\,d} = \frac{1}{\sqrt{3}}$$

and $$x = d/\sqrt{3}$$

Thus the ratio of the grating spaces of the 100, 110, and 111 planes are as

$$1 : \frac{1}{\sqrt{2}} : \frac{1}{\sqrt{3}}$$

From Fig. 5–10, it will be noted that each atom constitutes the junction of eight cubes and that each cube has eight atoms, one at each corner. If this structure is continued indefinitely in every direction then there will be one atom per cube, or half a molecule per cube. The mass of a cube is $Mm/2$ where M is the molecular weight of the salt and m the mass of the hydrogen atom. Hence

$$\frac{Mm}{2} = \rho d^3$$

and $$d = \sqrt[3]{\frac{Mm}{2\rho}}$$

where ρ is the crystal density.

Substituting for M, m, and ρ, $d = 2{\cdot}81 \times 10^{-8}$ cm.

In order to consider the effect of a crystal on X-rays, let a narrow beam of rays be incident upon a crystal plane as shown by

Fig. 5–12. As the wavefront AB meets the atoms in the plane, each atom may be considered as the origin of a spherical wave. Thus, when B reaches C, a spherical wavelet will arise, with C as centre,

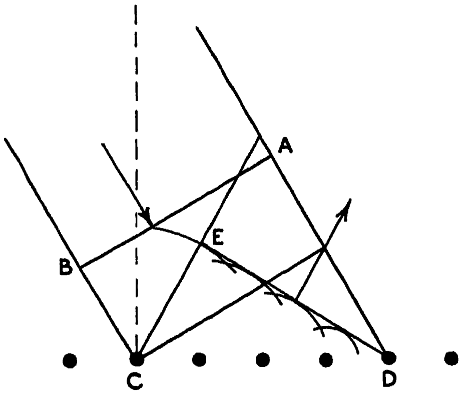


FIG. 5–12

and spread out with the velocity of light. When the wavefront reaches D the radius of any wavelet will be proportional to the distance of its atomic centre from D. Hence DE will be the reflected

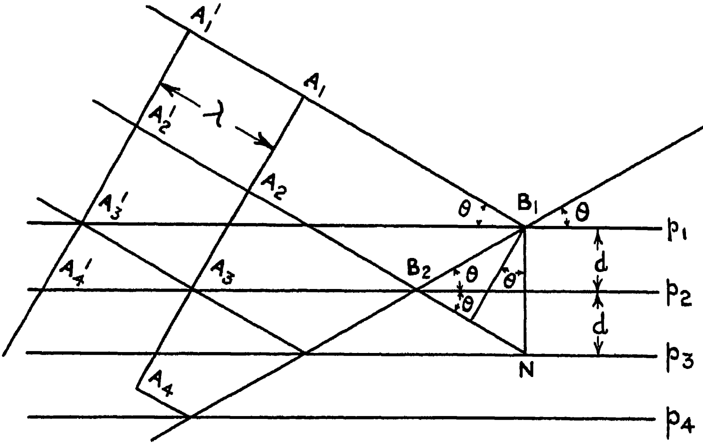


FIG. 5–13

wavefront moving in the direction indicated by the arrow, the angle of reflection being equal to the angle of incidence.

A crystal, of course, consists of a series of planes, and hence the effects of all planes must be considered with respect to the defraction of an X-ray beam. Let a series of planes, $p_1p_2p_3p_4$, be taken, each d cm. apart, as shown by Fig. 5–13. The line $A_1A_2A_3A_4$ represents a wavefront approaching the crystal planes, $A_1'A_2'A_3'A_4'$ being a second wavefront one wavelength, λ, behind the first. When the wavefront arrives at B_1 it is scattered by an atom in the p_1 plane to C. Similarly, the portion of the wavefront at A_2 strikes the p_2 plane at B_2 and is scattered along B_2B_1C. It is evident that the wavefront at A_2 is scattered from B_2 earlier than is the wavefront at A_1 from B_1. Also, there is a difference in the path lengths A_1B_1C and $A_2B_2B_1C$. The difference in these lengths is, of course, the difference in the distances A_1B_1 and $A_2B_2B_1$. From the geometry of the figure the difference of the path lengths is

$$\begin{aligned} & A_2B_2B_1 - A_1B_1 \\ &= (A_2N - B_2N + B_2B_1) - A_1B_1 \\ &= B_2B_1 - B_2N/2 \\ &= \left(\frac{B_1N}{\sin 2\theta} - \frac{B_1N}{\tan 2\theta}\right)\cos\theta \\ &= 2d\cos\theta\left(\frac{1}{\sin 2\theta} - \frac{1}{\tan 2\theta}\right) \\ &= 2d\sin\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5\text{–}7) \end{aligned}$$

Now, in order that the intensity of radiation along B_1C shall be a maximum, it is necessary that the rays reflected from the various planes shall all be in phase at B_1. This means that $2d \sin \theta$ must be a multiple of the wavelength λ. Hence for maximum reflexion we must have

$$n\lambda = 2d\sin\theta \quad . \quad . \quad . \quad . \quad (5\text{–}8)$$

If the intensity of the reflected radiation is measured by an ionization chamber, then, when a homogeneous beam of X-rays is directed on a crystal which is slowly rotated, a series of maxima and minima will be found for various values of θ corresponding to n. Thus, if $n = 1, 2, 3, 4$, etc., maximum intensity of ionization will occur for

$$\theta_1 = \text{arc sin } \lambda/2d$$

$$\theta_2 = \text{arc sin } \lambda/d$$

$$\theta_3 = \text{arc sin } 3\lambda/2d$$

etc.

W. H. BRAGG'S X-RAY SPECTROMETER

In order to measure the wavelengths of X-rays and also to study crystal structure, the X-ray spectrometer due to Bragg is generally employed, the principles of this instrument being shown by Fig. 5–14. X-rays from a suitable source are directed through the slits in the metal screens *A* and *B* on to the face of a crystal at *C*. The crystal is mounted upon a rotatable table, the angle of rotation being measured with the vernier *V*. After reflexion the X-rays pass through the slit *D* and then through a thin aluminium window into the ionization chamber *E*. The ionization chamber and the slit *D* can be rotated about the crystal axis, the angle of rotation being measured by the second vernier V_1. The chamber *E* consists of a cylindrical lead box containing an insulated electrode, the latter being connected to a sensitive electrometer. The chamber is raised to a potential of about 100 volts by a battery, and the conductivity of the gas in the chamber is measured in the manner described on page 59.

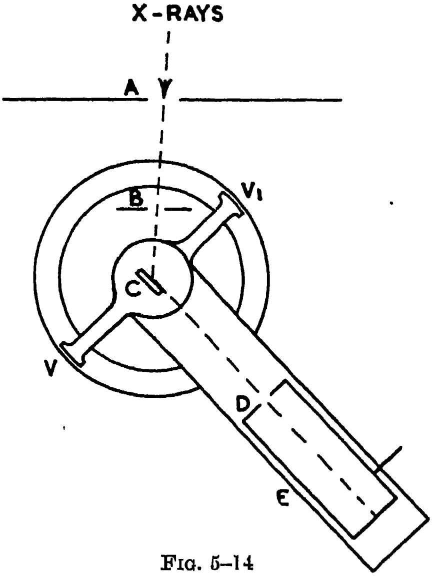


FIG. 5–14

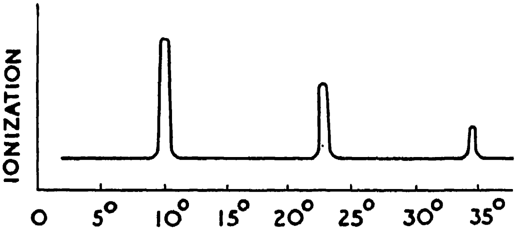


FIG. 5–15

In order to operate the spectrometer, X-rays are directed on to the crystal through the slits *A* and *B*, the angle between the crystal and the incident rays being one-half of that between the incident rays and the axis of the ionization chamber. The crystal and chamber are rotated and the ionization measured for various values of θ. A curve may then be plotted similar to that shown by Fig. 5–15, the various maxima, A_1, A_2, A_3, etc., being termed first, second, and third order reflexions. For X-rays derived from rhodium and reflected from a rock salt crystal A_1, A_2, and A_3, occur at angles which are approximately 11·8°, 23·5°, and 36°. Taking the sines of these angles we have

$$\sin 11{\cdot}8^\circ : \sin 23{\cdot}5^\circ : \sin 36^\circ = 0{\cdot}204 : 0{\cdot}04 : 0{\cdot}63 = 1 : 2 : 3$$

which is in accordance with (5–8). For the 100 planes in rock salt, $d = 2{\cdot}81 \times 10^{-8}$ cm. Hence for $\theta = 11{\cdot}8^\circ$

$$\begin{aligned}\lambda &= 2 \times 2{\cdot}81 \times 10^{-8} \times 0{\cdot}204\\ &= 1{\cdot}15 \times 10^{-8} \text{ cm.}\end{aligned}$$

X-rays and Atomic Number

Following the development of the X-ray spectrometer, it was found that the frequencies of the characteristic X-rays of the elements are simply related to their atomic numbers. Thus, if the square root of the frequency of, say, the K series is plotted against atomic number a straight line is obtained. Hence if f is the frequency, we have

$$f_{K\alpha} = aR(Z-b)^2$$

Also

$$f_{L\alpha} = cR(Z-d)^2$$

where a is found to equal 3/4, c, 5/36, and R is equal to the Rydberg constant.

Now

$$\frac{3}{4} = \frac{1}{1^2} - \frac{1}{2^2}$$

and

$$\frac{5}{36} = \frac{1}{2^2} - \frac{1}{3^2}$$

Substituting, we have

$$f_{K\alpha} = R(Z-b)^2\left(\frac{1}{1^2} - \frac{1}{2^2}\right) \qquad . \quad . \quad (5\text{–}9)$$

$$f_{L\alpha} = R(Z-d)^2\left(\frac{1}{2^2} - \frac{1}{3^2}\right) \qquad . \quad . \quad (5\text{–}10)$$

For the K_α series, b is found to be 1, while for the L series, d is 7·4.

Now, according to Bohr's theory of the hydrogen spectrum the frequencies are given by

$$f = RZ^2\left(\frac{1}{n_1{}^2} - \frac{1}{n_2{}^2}\right)$$

Comparing this with (5–9) and (5–10) it appears that Z in the latter two equations is the atomic number of the element concerned, b and d occurring due to the presence of other electrons in the K and L shells. Thus the effective value of the nuclear charge for the K series is $(Z - 1)$ and for the L series $(Z - 7{\cdot}4)$. This indicates that one electron of the two electrons of the K shell is radiating while the other neutralizes one of the positive charges of the nucleus. In the case of the L series, the value $(Z - 7{\cdot}4)$ indicates that about 7 to 8 electrons in the K and L shells neutralize the same number of positive charges of the nucleus.

From the foregoing it is evident that the relation between frequency and atomic number tends to confirm the theory of atomic structure outlined in Chapter I.

X-ray Tubes

As stated previously, early X-ray tubes were of the type indicated by Fig. 5–1, this being termed a gas tube. This means that gas is introduced into the tube in order to produce electrons from the cathode by positive ion bombardment. The gas tube has several disadvantages, among which are clean-up* and the impossibility of varying independently of each other the intensity and wavelength of the X-rays. Nevertheless, good results may be obtained with gas tubes, as is shown by Figs. 5–17, 5–18, 5–19, and 5–20, which were obtained by the author with a tube of this type.

The majority of tubes now in use produce the electrons thermionically. The cathode works at saturation and hence variation of the anode (or target) voltage will vary the wavelength of the rays without altering their intensity. Again, variation of the cathode heating current (and hence cathode temperature) will vary the intensity of the rays without affecting their wavelength. Tubes of this type will, of course, act as rectifiers, and hence may be employed with a.c. supplies without further rectifying equipment. However, in this case only comparatively low target voltages may be employed, for otherwise the high temperature of the target focal spot may cause the tube to pass current during the inverse half-cycle. For voltages in excess of 100 kV it is usual to employ a separate rectifying circuit of the form shown by Fig. 13–14.

THE TARGET

In order to produce X-rays the electron stream is directed at a target. The latter consists of a piece of metal of high atomic number and melting-point embedded in a massive copper member. The purpose of the copper member is, of course, to cool the target. From the medical viewpoint it is desirable that the focus of the tube and the time of exposure of a subject shall be as small as possible. A fine-focus tube, i.e. one having a small focal spot, results in radiographs possessing high definition, while a short exposure circumvents the effects of any motion on the part of the subject. Unfortunately, these two requirements are in opposition to each other. If the tube loading is increased with the object of shortening the exposure time the increased rate of energy dissipation at a small focal spot may result in melting and destruction of the target. This will be readily appreciated when it is stated that the efficiency of an X-ray tube is only about 1 to 2 per cent.

---

* See p. 625.

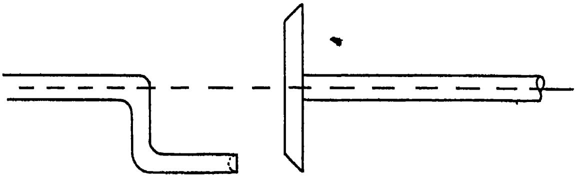


FIG. 5-16

In order to prevent damage to the target, various methods are employed for maintaining this at a relatively low temperature. One method has already been stated as embedding the target material (usually tungsten) in a block of copper. In addition to this, cooling fins may be added to an extension of the anode external to the tube. A further method is to pass a continuous stream of water up a hollow anode stem and across the under surface of the target.

If the area of the focal spot is increased in order to increase the target life, not only is it necessary to increase the exposure time, but the radiographs lose definition. These shortcomings have been overcome to some extent by the employment of what are known as line focus tubes. Such tubes employ a long cylindrical spiral filament of very small diameter and produce a focal spot rectangular in form, the length of which is about ten times the breadth.

A method of obtaining a focal spot of minimum area in a high-energy tube is to employ a target rotating at high speed. The target may either be driven mechanically through vacuum-tight joints or by the rotor of an induction motor enclosed within the tube. The motor stator is fitted outside and over the tube, and drives the rotor by virtue of its rotating field. The anode consists of a disc of tungsten with a bevelled edge as shown by Fig. 5-16, the electron bombardment being concentrated on the bevel. Thus, from the heating viewpoint the target length is $2\pi r$, where r is the anode radius.

Applications of X-rays

Apart from the evident applications of X-rays to the study of the structure of matter, numerous other applications exist. Only the shortest account of these can be given here, and for further information the reader is referred to works dealing specifically with this subject.*

RADIOGRAPHY

As already stated, the fact that X-rays are capable of penetrating solid bodies is the basis of the science of radiography. The degree of penetration depends on the "denseness" of the body, the "denser" the body the less the penetration. Hence, if a substance of varying denseness is placed over a photographic plate and then irradiated by an X-ray beam, a shadow picture may be obtained of the interior of the substance. This arrangement is, of course, analogous to that of obtaining records of the structure of small bodies under the microscope by transmitted light, or, again, corresponds to the practice of obtaining a "skeleton" picture of a leaf by exposing it to sunlight while in close contact with a sheet of sensitized photographic paper. The importance of radiography in medical diagnosis is, of course, so important and obvious that it is universally known. Practically any part of the body where there is sufficient variation in denseness may be studied, either by means of photography or a fluorescent screen. The tube voltage and current and the exposure

TABLE 5-3

LOCALITY	kV	MILLIAMP	SECONDS	DISTANCE INCHES
Chest—				
Posterior-anterior	88	500	—	72
Lateral	94	400	—	60
Gastro-intestinal tract—				
Stomach	75	300	0·25	30
Colon	75	300	0·5	30
Spine	20	100	2·0	40
Head	70	100	2·0	25
Extremities	45	200	0·05	40

---

* *Applied X-rays*, G. L. Clarke (McGraw-Hill). *Metallurgical and Industrial Radiology*, K. S. Low (Pitman).

time naturally depend on the part of the body undergoing examination and, to some extent, on the age of the patient. Some typical values of these quantities are given by Table 5-3.

Figs. 5-17 and 5-18, respectively, show a male adult's hand and a rat, and were obtained by the author with a gas tube operated by an induction coil. The figures given by Table 5-3, of course, refer to hot cathode tubes.

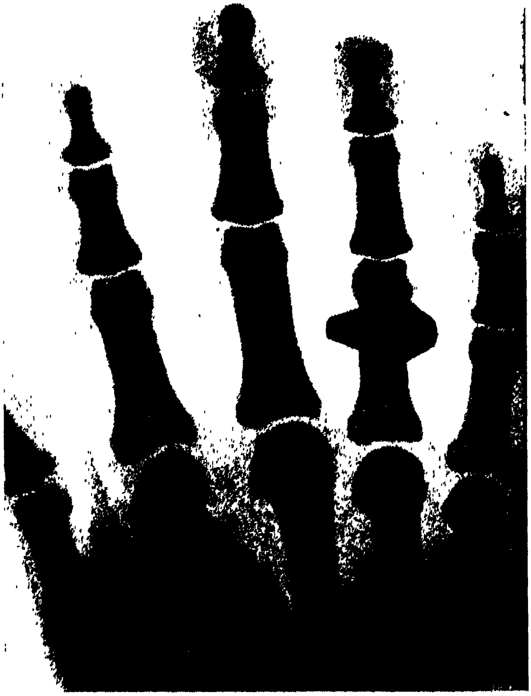


FIG. 5-17

INDUSTRIAL RADIOGRAPHY

In addition to animal bodies, most objects may be radiographed with the purpose of studying their structure and determining any defect or the presence of unhomogeneity. A particular application is the study of metals, particularly castings. In this case a variety of defects may be detected without affecting or destroying the specimens. The various defects that may be radiographically detected are as follows—

(a) Blow-holes, or gaseous occlusions.
(b) Inclusions, such as sand, slag, oxide, etc.
(c) Porosity.
(d) Shrinkage cavities.
(e) Cracks.
(f) Segregation.

FIG. 5–18

In the case of blow-holes, porosity, and cracks, regions exist within the metal having little or no absorbing power. Hence corresponding to such regions the radiograph will show localities which are darker than the surrounding unaffected parts of the metal. If x is the thickness of a sound specimen, then the transmitted intensity may be written $I_a = I_0\varepsilon^{-\mu x}$. For a specimen containing a cavity of thickness d, $I_b = I_0\varepsilon^{-\mu(x-d)}$. Hence

$$\frac{I_a}{I_b} = \frac{\varepsilon^{-\mu x}}{\varepsilon^{-\mu(x-d)}} = \varepsilon^{-\mu d}$$

From this result it is evident that, if d is small, difficulty will be experienced in recognizing the presence of blow-holes, cracks, etc.

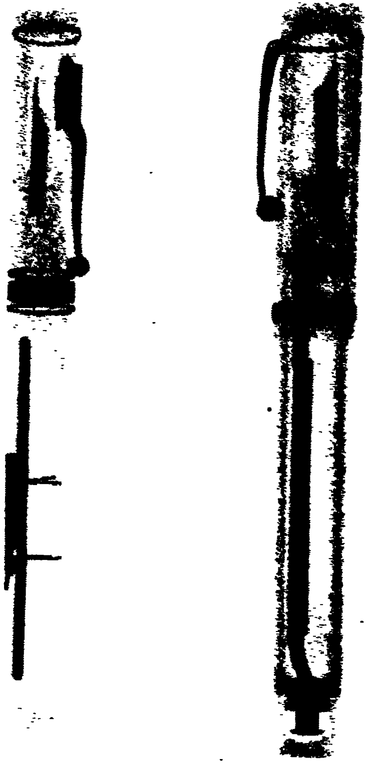


FIG. 5–19

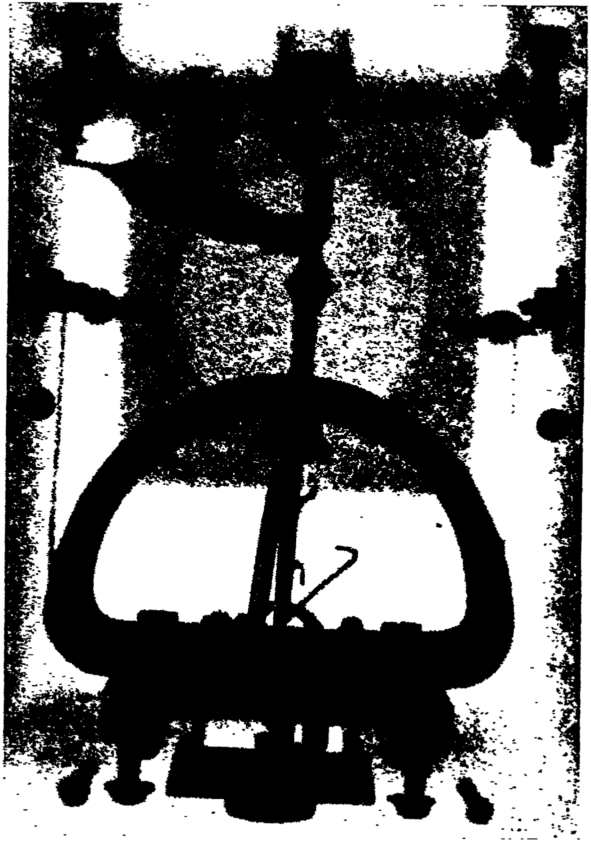


FIG. 5–20

In practice, for aluminium, iron, or copper, the dimension of the cavity in the direction of radiation must be not less than 5 per cent of the total thickness of the metal in that direction.

Numerous other industrial applications exist, in addition to the examination of metals. Among these may be mentioned the study of the position of metal inserts in bakelite and other mouldings. Fig. 5–19 reveals the internal structure of two well-known makes of fountain pens. This radiograph was obtained by the author with the gas tube already mentioned. A further application is the inspection of instrument assemblies, etc., enclosed within some case which it would, perhaps, be inconvenient to remove. In this connexion the radiograph of Fig. 5–20 may be considered, which shows a D'Arsonval galvanometer enclosed within a wooden box.

For further applications and appropriate technique the reader is referred to the works already mentioned.

BIBLIOGRAPHY

Six Papers on X-Rays, *J.A.I.E.E.*, Dec., 1945, pp. 423–444.

The Growth of Industrial Radiology, L. Mullins, "Electronic Engineering," Nov., 1945, p. 760.

X-Ray Diffraction in Inorganic Chemistry, Metallurgy and Minerology, H. Lipson, "Nature," Vol. 157, p. 124.

The Cystalline State, W. H. Bragg and W. L. Bragg, (Bell).

A Survey of X-Rays in Engineering and Industry, V. E. Pullin, *J.I E.E.*, Pt. I, June, 1945.

CHAPTER VI

ELECTRON OPTICS

THE ability of electric and magnetic fields to direct the paths of electrons and the fact that the latter are able to excite luminescence or affect a photographic film indicate the possibility of an analogy between certain forms of electron behaviour and geometrical optics. Thus a field may act upon a beam of electrons in a similar manner to a lens on a beam of light and, because of this formal analogy, such phenomena are termed *geometrical electron optics* or, more shortly, *electron optics*. Electron optics has important applications in electronic devices, such as cathode-ray tubes and electron microscopes, where the surface concentration of electrons emitted from one plane, such as a cathode, must be reproduced on a distance plane with a magnification either larger or smaller than unity. In the case of the cathode-ray tube the magnification is usually smaller than unity in order that a small brilliant spot may be formed on a fluorescent screen or a sharp fine trace on a photographic film. In an electron microscope the magnification is desired larger than unity for exactly the same reason as it is in an optical microscope.

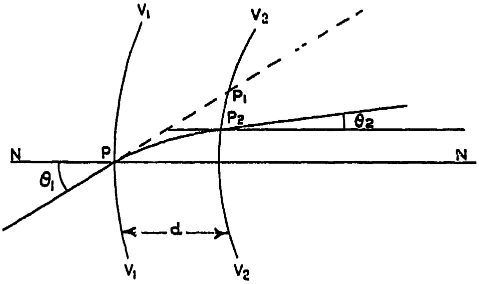


FIG. 6 1

Refraction of Electrons

In order to illustrate the refracting power of an electric field on an electron beam, Fig. 6–1 may be considered. Here V_1V_1 and V_2V_2 represent two equipotential surfaces at an infinitely small distance apart. The potentials V_1 and V_2 are assumed to be positive to, and measured from, some point at which electrons subsequently passing through the surfaces have zero velocity. At the point P

of the surface V_1V_1, at which an electron passes through, the electron velocity may be resolved into two components $v_1 \sin \theta_1$ and $v_1 \cos \theta_1$, the former being perpendicular to the normal N, and the latter parallel to N. When the electron enters the interspace d it will tend to follow the direction of the electrostatic field, i.e. along the normal N. Thus, the velocity $v_1 \cos \theta_1$ will be increased while $v_1 \sin \theta_1$ will remain unchanged. It is evident that the trajectory within the interspace will be bent or refracted towards the normal, taking the path PP_2 instead of PP_1.

Refractive Index

The path of the electron within d is not straight as is a light ray in a normal optical system, but curved as shown. This, of course, is due to the continually increasing velocity of the normal component of the electron. The path of the electron may be determined in the following manner. The acceleration imparted to the electron along the normal between V_1V_1 and V_2V_2 is

$$\frac{e(V_1 - V_2)}{md}$$

As the normal velocity at V_1 is $v_1 \cos \theta_1$, the velocity at any time t after leaving V_1 is

$$v_1 \cos \theta_1 + e(V_2 - V_1)t/md$$

Hence the distance travelled by an electron along the x axis in time t is

$$x = v_1 \cos \theta_1 t + e(V_2 - V_1)t^2/2md$$

In a similar manner

$$y = v_1 \sin\theta_1 t$$

Eliminating t

$$x = \frac{e(V_2 - V_1)}{2mdv_1^2 \sin^2 \theta_1} y^2 + \cos \theta_1 y$$

which shows the electron path between V_1 and V_2 to be parabolic.

The kinetic energy of an electron at V_1 is

$$\tfrac{1}{2}mv_1^2 = eV_1$$

and

$$v_1 = \sqrt{2eV_1/m} \quad . \quad . \quad . \quad (6\text{-}1)$$

Similarly at V_2

$$v_2 = \sqrt{2eV_2/m} \quad . \quad . \quad . \quad (6\text{-}2)$$

Now the velocity component perpendicular to N does not change.

Hence $$v_1 \sin \theta_1 = v_2 \sin \theta_2$$

and from (6–1) and (6–2)

$$\frac{v_2}{v_1} = \frac{\sin \theta_1}{\sin \theta_2} = \frac{\sqrt{V_2}}{\sqrt{V_1}} \quad . \quad . \quad . \quad (6\text{–}3)$$

The angles θ_1 and θ_2 are analogous to the angles of incidence and refraction in optics, and if $\sqrt{V_2}$ and $\sqrt{V_1}$ are considered as being analogous to the refractive indices of two adjoining media traversed by a light ray, then (6–3) is equivalent to Snell's Law.

For various reasons the analogy between geometrical and electron optics is by no means perfect. For example, generally speaking, optical systems are characterized by sudden changes in refractive index from one medium to another. Thus, in passing from air to glass, the refractive index suddenly changes from 1·0 to 1·5, and vice versa if the process is reversed. In contrast with this the electric fields established by the charges on the electrodes in an electron optical system result in potentials which vary continuously and not by discontinuous changes. The result is that such systems present a continuously variable refracting medium for electrons. The corresponding optical system should have the refractive index continuously varying in such a manner that $\mu \propto \sqrt{V}$ where μ is the refractive index. A further complication which is sometimes of importance is the modification of the field distribution by the presence of space changes. The influence of the latter may be appreciable in the vicinity of the cathode, but can usually be neglected in other regions.

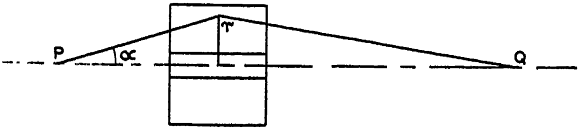


FIG. 6–2

A method of concentrating and focusing an electron beam is by means of radial field produced in the manner indicated by Fig. 6–2. Here, the electrodes take the form of two concentric cylinders of widely differing radii between which a potential difference V is maintained. A beam of electrons is produced at a point P, and it is desired to focus this at a point Q. If r_2 and r_1 are, respectively, the radii of the outer and inner electrodes, then the field at any point r between the electrodes is

$$X = \frac{V}{r \log \frac{r_2}{r_1}}$$

Consequently an electron leaving P at an angle α and entering the field at a radius r will experience a radial force

$$Xe = \frac{Ve}{r \log \frac{r_2}{r_1}}$$

This will result in a radial acceleration

$$\frac{d^2r}{dt^2} = \frac{Ve}{mr \log \frac{r_2}{r_1}}$$

or

$$\frac{dv}{dt} = \frac{Ve}{mr \log \frac{r_2}{r_1}}$$

If the variation in r is inappreciable during the passage of the electron through the field, the last expression may be integrated. Thus

$$\int_{v_1}^{v_2} dv = \int_{t_1}^{t_2} \frac{Ve}{mr \log \frac{r_2}{r_1}} dt$$

or

$$v_2 - v_1 = \frac{Ve}{mr \log \frac{r_2}{r_1}} (t_2 - t_1)$$

where v_2 and v_1 are, respectively, the radial velocities of the electron at the entry and exit of the field. Now $(t_2 - t_1)$ is the time of passage through the field and, if l is the length of the field and v_0 the axial velocity,

$$v_2 - v_1 = \frac{Ve}{mr \log \frac{r_2}{r_1}} \cdot \frac{l}{v_0} \quad . \quad . \quad . \quad (6\text{–}4)$$

Now the initial radial velocity v_1 and that produced by the field are opposite in direction. Thus if $v_2 < v_1$ the beam is still diverging after leaving the field, but less so than when entering.

If $v_2 = v_1$ an electron after leaving the field will travel parallel to the axis. For $v_2 > v_1$ the radial velocity is reversed and ultimately the electron will pass through a point on the axis. From (6–4) it will be noted that the change in radial velocity experienced by an electron in passing through the field varies inversely as the radius of the electron within the cylinder, and, consequently, inversely as the initial angle of divergence. Hence, electrons with different angles of divergence will experience different changes of radial velocity and will be brought to a focus at different positions on the axis. This is an effect analogous to spherical aberration in geometrical optics.

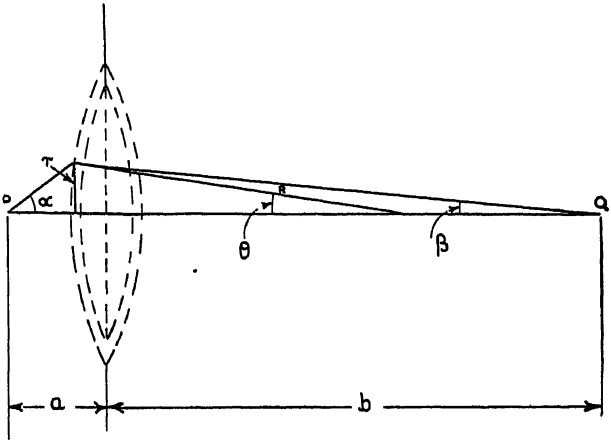


FIG. 6–3

In order to obviate the foregoing defect, it is evident from (6–4) that in order to bring all the electrons entering the field at different radii to a focus at Q, the change in radial velocity must be proportional to r instead of inversely proportional to this quantity. This, of course, means that the radial force acting on an electron should vary directly as its radial distance from the axis. In practice, radial fields possessing this property are difficult to produce from electrode arrangements suitable for electron optical systems. Hence systems have been devised possessing both axial and radial components, the latter having a field strength which varies directly as the distance from the axis.

As an example of the foregoing, the lens of Fig. 6–3 may be considered. This consists of spherical surfaces, formed of fine wire mesh, between which a potential difference is maintained. The latter produces a field which is approximately normal to the two concentric surfaces. Thus, an electron with an initial angle of divergence α will experience a radial force proportional to sin θ or r/R. If the curvature of the lens is small, $\sin\theta \doteqdot \theta$, and thus the radial force varies as θ. In addition to the radial force is an axial force proportional to cos θ. However, if the lens dimensions are small compared with R, this force is approximately independent of θ and constant. Moreover, the axial forces are equal and opposite in the two sides of the lens, and hence the net change in the axial velocity of an electron is zero. As the radial force varies as the distance of any electron from the axis, it is evident that the correct conditions exist for bringing all electrons to a focus at some point, such as Q.

FOCAL LENGTH

The focal length of the lens of Fig. 6–3 may be found in the following manner. The radial force on an electron at any radius r is given by

$$eX_r = e\frac{(V_2 - V_1)}{d}\cdot\frac{r}{R} = m\frac{dv}{dt} \quad . \quad . \quad (6\text{–}5)$$

where X_r is the radial component of the field, V_2 and V_1 are, respectively, the inner and outer potentials of the lens surfaces, d is the distance between these surfaces measured along the normal, and v is the radial velocity. Integrating (6–5) on the assumption that the change in r is inappreciable during the passage of an electron,

$$\frac{e}{m}\cdot\frac{V_2 - V_1}{d}\cdot\frac{r}{R}(t_2 - t_1) = v_2 - v_1 \quad . \quad . \quad (6\text{–}6)$$

Where $(t_2 - t_1)$ is the time for which an electron is under the influence of X_r and v_1 and v_2 are, respectively, the radial velocities at the moment of entering and leaving the lens. Providing that the lens is thin, i.e. R large, the axial distance travelled by an electron while under the influence of X_r is approximately equal to $2d$. Hence if v_0 is the axial velocity corresponding to V_1, $(t_2 - t_1) = 2d/v_0$, and (6–6) becomes

$$\frac{2e}{m}\cdot\frac{V_2 - V_1}{v_0}\cdot\frac{r}{R} = v_2 - v_1 \quad . \quad . \quad (6\text{–}7)$$

If an electron is incident upon the left-hand surface of the lens with an axial velocity only, the electron will ultimately pass through the lens focal point. Multiplying top and bottom of (6–7) by v_0 and writing

$$v_0^2 = 2eV_1/m$$

there results
$$\frac{v_2}{v_0} = \frac{r(V_2 - V_1)}{RV_1}$$

But, from Fig. 6–4, $v_2/v_0 = \tan\phi = r/f$, and thus

$$f = \frac{R}{\frac{V_2}{V_1} - 1} \quad . \quad . \quad . \quad . \quad (6\text{–}8)$$

From this expression it will be noted that if $V_2 > V_1$ the lens is converging, whereas if $V_2 < V_1$ the lens is diverging. It is evident that the focal length of the lens can be continuously varied over a large range of values merely by varying the ratio V_2/V_1. Thus, electron lenses possess much greater flexibility than those of the ordinary optical type.

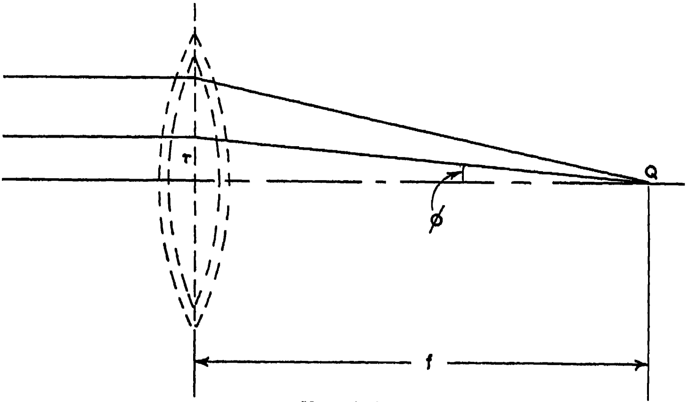


FIG. 6–4

In practice, a lens formed of wire mesh or gauze possesses several disadvantages which tend to preclude its use. Such are absorption losses, secondary emission, heating effects, obstruction, and electron deflexion by the wires of the mesh. However, practical electrode arrangements may be effected which simulate the equipotential curved surfaces of Fig. 6–3, this being the principle of all focusing systems. The disposition of the electrodes is such as to cause fringe fields to exist, which, of course, are accompanied by curved equipotential surfaces, the latter, when suitably curved, constituting an electron lens.

The Single-diaphragm Aperture Lens

An aperture lens consists of a plane electrode with a small circular hole at its centre, the plane being held at a positive potential. Such an arrangement exhibits a focusing property by imparting a radial velocity component to electrons passing through the aperture. The velocity component is, of course, due to a radial field component, the latter being produced by the non-uniformity of the field distribution in the vicinity of the aperture. To investigate the lens it is evident that the radial component of the field must first be determined.

As the system is cylindrically symmetrical, the axis of symmetry is the x-axis perpendicular to the plane and passing through the centre of the aperture. Consider a small cylindrical volume concentric with the axis, as shown by Fig. 6–5, this volume being close to the aperture but not actually within the plane. The field within the volume-element is symmetrically directed about the x-axis, i.e. for given values of x and r it does not vary with angular rotation about x. Referring to Fig. 6–6, which shows the distribution of the equipotential surfaces, it will be noted that these are similar in form to the lens surfaces of Fig. 6–3. Thus it is possible to make a similar assumption as before, i.e. providing the aperture is small compared with the radii of curvature of the equipotential surfaces, it may be assumed that the axial component of the field is independent of r.

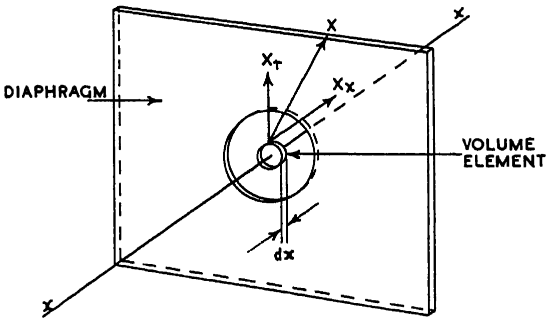


Fig. 6–5

If X_x is the field strength at the left-hand face of the volume-element, then the total flux entering this face is $\pi r^2 X_x$ and that leaving the right-hand face $\left(X_x + \frac{dX_x}{dx}\,dx\right)\pi r^2$. The flux leaving the circumferential face of the element is $2\pi r dx X_r$ and hence the net flux leaving the element is

$$\left(X_x + \frac{dX_x}{dx}\,dx\right)\pi r^2 + 2\pi r dx X_r - \pi r^2 X_x \quad . \qquad (6\text{–}9)$$

which, providing the space-charge within the volume-element is negligible, is equal to zero. Assuming this to be the case, (6–9) reduces to

$$X_r = -\frac{r}{2}\frac{dX_x}{dx}$$

It is immediately apparent that we have the necessary conditions for focusing, i.e. the radial electrostatic force varies directly as the radius. It is also clear that focusing cannot occur unless $dX_x/dx \neq 0$. In order that electrons which are parallel to or diverging from x may be directed towards the axis, it is necessary that X_r is directed away from the axis because electrons tend to travel in the opposite direction to that of electrostatic lines of force. This means, of course, that dX_x/dx must be negative.

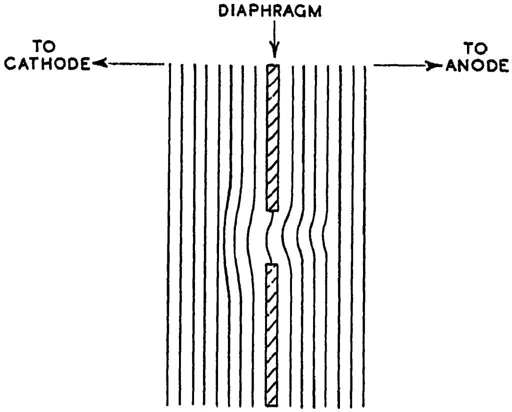


Fig. 6–6

Let an electron now pass through the aperture parallel to the axis, and at a distance r from it. If the lens is thin and the electron velocity high, it may be assumed that the change in r in passing through is negligible. The radial force acting on the electron is

$$e \cdot \frac{r}{2}\frac{dX_x}{dx} = m\frac{dv}{dt}$$

where v is the radial velocity. Multiplying both sides by dx we may write

$$m\frac{dv}{dt}\,dx = e\frac{r}{2}\,dX_x$$

or

$$m\,dv\,v_x = e\frac{r}{2}\,dX_x$$

where v_x is the axial velocity.

Integrating

$$\int_{v_{r1}}^{v_{r2}} dv = \frac{1}{2}\frac{e}{m}\int_{X_{x1}}^{X_{x2}} r\frac{dX_x}{v_x} \quad . \quad . \quad . \quad (6\text{–}10)$$

where v_{r2} and v_{r1} are, respectively, the radial velocities before and after passing through the lens, and X_{x2} and X_{x1} are the corresponding values of the electric field strength. If v_x is taken close to the lens it may be considered as being that corresponding to lens potential V, i.e. $v_{x0}^2 = 2eV/m$. Substituting in (6–10) we have

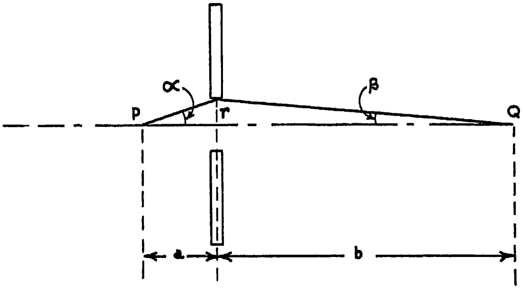


FIG. 6–7

$$v_{r2} - v_{r1} = \frac{r v_{x0}}{4V}(X_{x2} - X_{x1})$$

Referring to Fig. 6–7, if an electron originates at P and after passing through the lens cuts the axis at Q,

$$\tan\alpha = v_{r1}/v_{x0} = r/a$$

$$\tan\beta = -\,v_{r2}/v_{x0} = r/b$$

and

$$-\frac{r}{b} - \frac{r}{a} = \frac{r}{4V}(X_{x2} - X_{x1})$$

or

$$\frac{1}{b} + \frac{1}{a} = \frac{X_{x1} - X_{x2}}{4V}$$

which is analogous to the well-known relation in optics

$$\frac{1}{v}+\frac{1}{u}=\frac{1}{f}$$

Hence in the electron-optical case

$$f=\frac{4V}{X_{x1}-X_{x2}}$$

where X_{x1} and X_{x2} are, respectively, the axial field intensities at *a* and *b*. It is evident that if $X_{x1} > X_{x2}$ the lens is converging, whereas if $X_{x1} < X_{x2}$ the lens is diverging.

The Three-diaphragm Aperture Lens

A lens somewhat similar to the foregoing one is shown by Fig. 6–8. It will be seen to consist of three diaphragms with circular

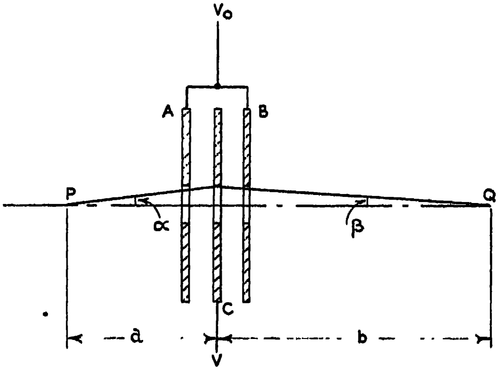


FIG. 6–8

centrally placed apertures, *A* and *B* being connected and maintained at the same potential as the anode responsible for accelerating the electrons in the cathode-anode space. Because *A* is at anode potential there is no force on electrons between *P* and *A*. Between *A* and *B* is a third diaphragm *C*, maintained at a potential V_1 such that $V_1 > V_0$. If the directions of the field lines between the diaphragms are considered, it will be found that the axial components on opposite sides of *C* are equal and opposite, while the radial components are equal and similar. Hence the net change in axial velocity of an electron passing through the lens is zero, while the radial components produce a radial force on the electrons, urging them towards the axis. If a small axial cylindrical volume-element is considered, as with the previous lens, it will again be found that

$$X_r = -\frac{r}{2}\frac{dX_x}{dx}$$

which indicates that the system will produce point focusing. Again, we have

$$m\frac{dv_r}{dt} = eX_r = \frac{er}{2}\frac{dX_x}{dx}$$

$$\int_{v_{r1}}^{v_{r2}} dv_r = \frac{er}{2m}\int_{-\infty}^{\infty}\frac{dX_x}{dx}\,dt$$

and $$\frac{v_{r2}}{v_{x0}} - \frac{v_{r1}}{v_{x0}} = \frac{er}{2mv_{x0}}\int_{-\infty}^{\infty}\frac{dX_x}{dx}\,dt \quad . \quad . \quad . \quad (6\text{–}11)$$

where v_{x0} is the electron velocity corresponding to V_0. But

$$v_{r1}/v_{x0} = r/a = \tan\alpha$$
$$v_{r2}/v_{x0} = -r/b = -\tan\beta$$

and substituting in (6–11)

$$\frac{1}{b} + \frac{1}{a} = -\frac{e}{m}\frac{1}{2v_{x0}}\int_{-\infty}^{\infty}\frac{dX_x}{dx}\,dt = \frac{1}{f}$$

which again is equivalent to the law in geometrical optics. Putting $dx/dt = v_x$,

$$\frac{1}{f} = -\frac{e}{m}\frac{1}{2v_{x0}}\int_{-\infty}^{\infty}\frac{1}{v_x}\frac{dX_x}{dx}\,dx \quad . \quad . \quad (6\text{–}12)$$

If V is the potential at any point within the lens then

$$\tfrac{1}{2}mv_x^2 = eV_0 + eV$$

from which $$v_x = \sqrt{2\frac{e}{m}(V_0 + V)}$$

where V is measured relative to V_0.

Also $$v_{x0} = \sqrt{2\frac{e}{m}V_0} \text{ and } \frac{dX_x}{dx} = -\frac{d^2V}{dx^2}$$

Substituting in (6–12),

$$\frac{1}{f} = \frac{1}{4V_0^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{1}{(V_0 + V)^{\frac{1}{2}}}\frac{d^2V}{dx^2}\,dx$$

Integrating by parts

$$\frac{1}{f} = \frac{1}{4V_0^{\frac{1}{2}}}\left[\frac{1}{(V_0 + V)^{\frac{1}{2}}}\frac{dV}{dx}\right]_{-\infty}^{\infty} + \frac{1}{8V_0^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{1}{(V_0 + V)^{\frac{3}{2}}}\left(\frac{dV}{dx}\right)^2 dx$$

As the forces acting on an electron due to V are negligible on either side of the lens, dV/dx vanishes at the limits $\pm\,\infty$, the latter only indicating the points at which the field is free from distortion. Consequently we have

$$\frac{1}{f} = \frac{1}{8V_0^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{1}{(V_0 + V)^{\frac{3}{2}}}\left(\frac{dV}{dx}\right)^2 dx \qquad . \quad . \quad . \quad (6\text{–}13)$$

which can only be determined if V can be expressed as a function of x and also in a form which can be readily integrated. In practice, V is usually determined by an experimental method described on page 243 and (6–13) then integrated graphically.

In considering the potential distribution of the lens of Fig. 6–8, the distribution will tend to differ according to the value of r at which it is taken. As dX_x/dx may be written $-\,d^2V/dx^2$ the radial force can be expressed as

$$X_r = \frac{r}{2}\frac{d^2V}{dx^2} \qquad . \quad . \quad . \quad . \quad (6\text{–}14)$$

Thus, if V is a linear function of x, $d^2V/dx^2 = 0$ and there is no radial force. Again, should $V \propto -\sin\theta$, $d^2V/dx^2 \propto \sin\theta$. Certain distributions may be found, however, in which d^2V/dx^2 changes sign so that an electron is subjected to equal but opposite forces for equal distances through the lens. In such cases it might seem that no convergence would occur. However, when X_r is positive dX_x/dx is negative. This means that the axial velocity is slower than when X_r is negative and dX_x/dx positive. Hence electrons pass more slowly through convergent sections of the lens than through those having a divergent effect, with the net result that a converging action is always produced. This is true, whether the centre diaphragm is positive or negative, and thus the three-diaphragm lens is always converging. These results are actually implicit in (6–13), where it will be noted that the integral not only depends on the electrostatic force but also upon the axial velocity.

Two-diaphragm Lens

The two-diaphragm lens is similar to that of Fig. 6–8 and is illustrated by Fig. 6–9. Its mathematical treatment is exactly as for the three-diaphragm lens, but, unlike the latter, the electrons undergo a net change in axial velocity which has the effect of changing the image size. To understand this the magnification of the lens must be considered.

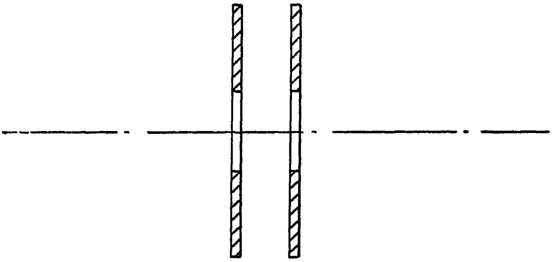


Fig. 6–9

Referring to Fig. 6–10, let V and $V + dV$ be two adjacent equipotential surfaces of an electrostatic lens. Consider an electron passing through P_1, the tangent to its direction of flight being $\tan \alpha$ and the angle between its direction and the normal to the equipotentials, θ. The force acting on the electron is directed along the normal and is equal to $eX = e\,dV/dl$. The component of this force perpendicular to the direction of flight is

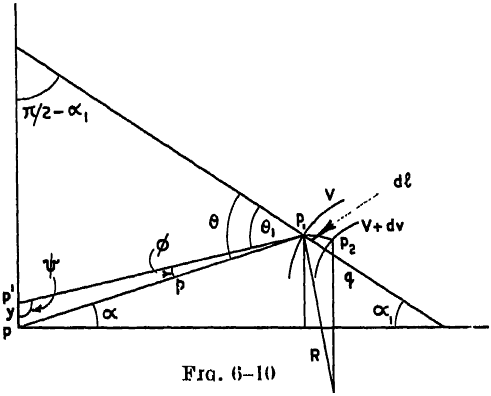


Fig. 6–10

$$e\frac{dV}{dl}\sin\theta$$

this being balanced by a centrifugal force equal to mv^2/R, R being the radius of curvature of the path of the electron at P_1. If v_0 is the electron velocity before entering the lens, then

$$\begin{aligned}\tfrac{1}{2}mv^2 &= \tfrac{1}{2}mv_0^2 + eV \\ &= eY\end{aligned}$$

where $Y = V + V_0$, V_0, of course, corresponding to v_0 and being related to the latter by $v_0 = \sqrt{2eV_0/m}$. Hence we have

$$\frac{mv^2}{R} = \frac{2eY}{R} = e\frac{dV}{dl}\sin\theta$$

and
$$\frac{1}{R} = \frac{dV}{dl}\frac{\sin\theta}{2Y} \quad . \quad . \quad . \quad . \quad . \quad (6\text{–}15)$$

In passing between the equipotential surfaces the electron is deflected through an angle $d\theta$, this angle being given by

$$d\theta = -P_1P_2/R$$

as θ decreases as V increases.

Replacing P_1P_2 by $dl/\cos\theta$

$$d\theta = -dl/R\cos\theta$$

and substituting for $1/R$, from (6–15)

$$d\theta = -\frac{dl\dfrac{dV}{dl}\sin\theta}{2Y\cos\theta}$$

Hence
$$\frac{d\theta}{\tan\theta} = -\frac{dV}{2Y} = -\frac{dV}{2(V + V_0)}$$

Integrating
$$\int\frac{d\theta}{\tan\theta} = -\frac{1}{2}\int\frac{dV}{(V + V_0)}$$

$$\log_e\sin\theta = -\tfrac{1}{2}\log_e(V + V_0) + \log K'$$

where K' is a constant, and

$$\sin\theta = K'Y^{-\frac{1}{2}}$$

or
$$\theta = K'Y^{-\frac{1}{2}} \text{ if } \theta \text{ is small} \quad . \quad . \quad . \quad (6\text{–}16)$$

Consider now a second electron passing through P_1 and originating at a point P' above the axis so that $PP' = y$. From the sine rule

$$\frac{y}{\sin\phi} = \frac{p}{\sin\psi}$$

But
$$\psi = (\pi/2 - \alpha_1) + \theta_1 = \pi/2 - (\alpha_1 - \theta_1)$$

and
$$\sin\psi = \cos(\alpha_1 - \theta_1)$$

Thus
$$y = p\,\frac{\sin \phi}{\sin \psi}$$
$$= p\,\frac{\sin (\theta - \theta_1)}{\cos (\alpha_1 - \theta_1)}$$

Again $p/\sin \alpha_1 = q/\sin \alpha$, and $p = q \sin \alpha_1/\sin \alpha$. Substituting for p
$$y = q\,\frac{\sin \alpha_1}{\sin \alpha}\,\frac{\sin (\theta - \theta_1)}{\cos (\alpha_1 - \theta_1)}$$

If $(\theta - \theta_1)$ and $(\alpha_1 - \theta_1)$ are small, then
$$y = q\,\frac{\sin \alpha_1}{\sin \alpha}(\theta - \theta_1)$$

Now, according to (6–16)
$$\theta = K_1' Y^{-\frac{1}{2}}$$
$$\theta_1 = K_2' Y^{-\frac{1}{2}}$$

and, therefore
$$y = q\,\frac{\sin \alpha_1}{\sin \alpha}\,K Y^{-\frac{1}{2}}$$

where K is constant. As $q \sin \alpha_1$ is a constant for all electrons passing through P_1, it follows that for all electrons passing through this point
$$y \sin \alpha Y^{\frac{1}{2}} = \text{constant}$$

which is equivalent to the Abbe's Sine Law in geometrical optics.

Linear Magnification

A pair of electrons originating from two points PP' of an object, will ultimately form two corresponding points of its image. Thus, if y_0 and y_i, respectively, represent the heights of the object and image, and $Y_0^{\frac{1}{2}}$ and $Y_i^{\frac{1}{2}}$ are, respectively, the potentials at the object and image planes, then
$$y_0 \sin \alpha Y_0^{\frac{1}{2}} = y_i \sin \beta Y_i^{\frac{1}{2}}$$

and the magnification is
$$m = \frac{y_i}{y_0} = \frac{\sin \alpha}{\sin \beta}\,\frac{Y_0^{\frac{1}{2}}}{Y_i^{\frac{1}{2}}}$$

Referring to Fig. 6–7, if α and β are small, $\sin \alpha$ and $\sin \beta$ are, respectively, approximately given by r/a and $-r/b$
$$m = -\frac{Y_0^{\frac{1}{2}}}{Y_i^{\frac{1}{2}}}\,\frac{b}{a}$$

or
$$\frac{Y_0^{\frac{1}{2}}}{Y_i^{\frac{1}{2}}}\,\frac{b}{a} \text{ (taking absolute values)}$$

Considering the three-diaphragm lens of Fig. 6–8, at P $Y_0 = V_0$. Also at Q, $Y_i = V_0$. Hence in this instance

$$m = \frac{b}{a}$$

which corresponds to the case in geometrical optics when the refractive indices in the object and image spaces are the same. For the two-diaphragm lens $Y_0 = V_0$ and $Y_i =$ (say) V_1. Hence

$$m = \sqrt{\frac{V_0}{V_1}}\frac{b}{a}$$

which is analogous to geometrical optics when the refractive indices in the object and image spaces are different. As $V_1 > V_0$ it follows that the magnification with a two-diaphragm lens is less than that of a three-diaphragm lens for the same values of V_0, V_1, a, and b. It follows that for oscillograph work (where a small spot is required) this is advantageous.

Determination of Potential Distribution

It is evident that in order to calculate the focal length of lenses such as those of Figs. 6–8 and 6–9, a knowledge of the potential distribution along the axis is necessary. Furthermore, if rays remote from the axis must be considered, then the potential distribution in these regions must also be known. As the theoretical determination of the distribution is often a difficult or impossible problem, an experimental method is usually employed. One method consists of the employment of what is known as the "electrolytic trough." With this method an enlarged model of the lens system under consideration is immersed in a trough containing an electrolyte, the dimensions of the trough being large compared with those of the model. To the various electrodes of the lens, alternating potentials are applied which are proportional to the d.c. potentials which would be used in the actual lens system. As the lens possesses cylindrical symmetry it is only necessary to construct a model of one-half of the system. This is then immersed in the trough so that the surface of the electrolyte is in the plane of symmetry. The potential at any point in the liquid is ascertained by means of a probe, the probe being connected through some null indicator, such as a telephone receiver, to the slider of a potentiometer. One end of the potentiometer is connected to an electrode of the system, and all potentials are measured relatively to this. When the potential of a position in the model system is equal to that of the potentiometer, setting the indicator will show this by an absence of deflexion, if an instrument, or an absence of sound, if a telephone.

Fig. 6-11 shows the potential distribution between two coaxial cylinders when maintained at a potential difference of 100 volts.

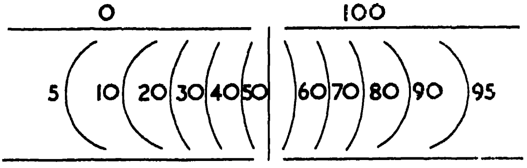
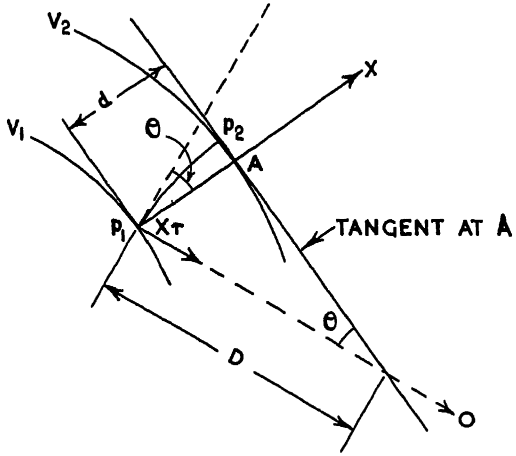


Fig. 6-11

Determination of Electron Trajectory

If it is desired to follow the trajectory of an electron from point to point in an electrostatic field of which the potential distribution is known, the following method may be employed. Referring to Fig. 6-12, let V_1 and V_2 be two adjacent equipotential surfaces which have been determined, say, by the electrolytic trough method. An electron passing through the point P_1 will experience an electrostatic force directed along the normal to the two surfaces. This force may be resolved into two components, one parallel to the



O = INSTANTANEOUS CENTRE OF CURVATURE

Fig. 6-12

path of the electron and the other perpendicular to this. The latter force will alter the direction of the electron, causing it to follow a curved course, the centre of curvature of which is in the direction of the perpendicular component. From Fig. 6–12 the value of this component is

$$X_r = X \sin \theta$$

and this will be balanced by a centrifugal force, due to the electron, of magnitude mv^2/R, where R is the radius of curvature of the trajectory. It follows that

$$\frac{mv^2}{R} = eX_r = eX \sin \theta$$

But

$$v = \sqrt{2eV/m}$$

where V is the instantaneous potential of the electron. Thus

$$\frac{2eV}{R} = eX \sin \theta$$

and

$$R = \frac{2V}{X \sin \theta}$$

Replacing X by $(V_2 - V_1)/d$ and $\sin \theta$ by d/D

$$R = \frac{2VD}{V_2 - V_1}$$

or, if V is taken as the mean of V_1 and V_2,

$$R = D\frac{V_1 + V_2}{V_2 - V_1}$$

From this result the centre of curvature may be found corresponding to any point such as P_1 and the electron trajectory approximated to by drawing a circular arc P_1P_2 with R as radius and O as centre. The process is then repeated from P_2 to the next equipotential, P_3, and so on. The actual trajectory is obtained by the smooth curve which connects all the elementary circular arcs.

Magnetic Focusing

An alternative method to electrostatic focusing is that employing a magnetic field. To introduce this method, let an electron be emitted from some source within a uniform magnetic field H. If the electron velocity is v, this may be resolved into two components $v \sin \theta$ and $v \cos \theta$, the former being perpendicular to H and the latter parallel to H. As shown on page 124, the component $v \sin \theta$ will cause the electron to describe a circular path of radius

$$r = \frac{mv \sin \theta}{He} \text{ cm.}$$

the time for one revolution being

$$t = \frac{2\pi m}{He}$$

Because of the axial component, $v \cos \theta$, the path of the electron is a helix of pitch

$$p = \frac{2\pi\, mv \cos \theta}{He}$$

As t is independent of θ, it follows that if a number of electrons are emitted from a point with different angles of emergence (θ) but equal axial velocities, they will all pass through another common point p distant from the first after a time t. Hence an image of the point of emergence will be formed at a distance p. Assuming the source is the anode aperture of a cathode-ray tube, it is thus possible to form an image of this if the screen is placed at a distance np from the source, n being an integer. It is to be noted that for image formation to be effective, p must be the same for all electrons. This means that all electrons must have the same axial velocities, and hence the same values of $v \cos \theta$. This will be approximately so, providing θ is sufficiently small for $\cos \theta$ to be essentially unity.. Fig. 6–13 shows an end view of the trajectories of electrons emitted from a point source on the axis of a tube immersed in a uniform axial magnetic field. It will be noted that the electron paths appear as circles, each passing through a common point on the axis. The various points, each separated from the other by a distance p, through which an electron passes, all lie on a straight line at a constant distance from the field axis. Hence the magnification is unity.

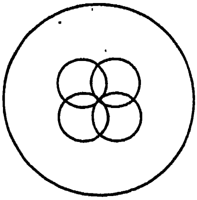


FIG. 6–13

SHORT SOLENOID FOCUSING

The above method of magnetic focusing necessitates a uniform field and, as applied to a cathode-ray tube, would require the field to be maintained along the entire length of the tube. Furthermore, on account of the lack of magnification, it is obvious that the system cannot be employed for an electron microscope. Because of these disadvantages short focusing coils are commonly employed, and the theory of this method of focusing will now be considered.

Considerations will be restricted to the field due to a short solenoid at points near the axis, as the electron beam radius is normally small compared with that of the solenoid. At any point x within the solenoid the radial component of the field will be assumed to be independent of r but to vary with x. The value of the radial component can be found in an exactly similar manner to that previously employed for finding the radial component of an electrostatic field. Thus the flux entering the left-hand face of a cylindrical volume-element concentric with the axis is

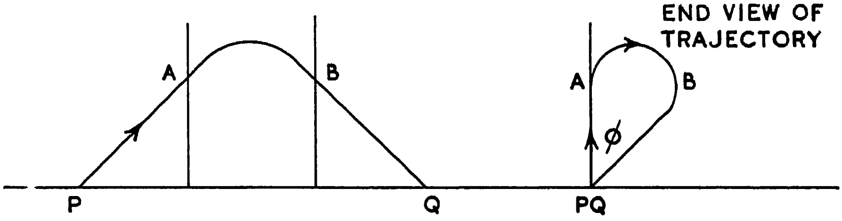


Fig. 6–14

$$\pi r^2 H_x$$

and that leaving the right-hand face

$$\pi r^2 \left(H_x + \frac{dH_x}{dx} dx \right)$$

The flux passing outwards across the periphery is

$$2\pi r \,.\, dx \,.\, H_r$$

and hence $\pi r^2 H_x = \pi r^2 \left(H_x + \frac{dH_x}{dx} dx \right) + 2\pi r \,.\, dx \,.\, H_r$

from which $H_r = -\frac{r}{2} \frac{dH_x}{dx}$

Let it now be assumed that an electron passing through a point P on the axis moves obliquely to the axis in the direction PA, as shown by Fig. 6–14. Assuming P to be outside the field, the electron will follow a linear path until it comes within the influence of the field. Within the field the electron comes under the influence of two opposing forces; one due to its axial velocity and the radial component of the field, and the other due to its radial velocity and the axial component of the field. These conditions are shown by the vector diagrams of Fig. 6–15. Considering the first force, this is equal to

$$ev_xH_r$$

and tends to rotate the electron anti-clockwise. The second force is

$$ev_rH_x$$

and tends to rotate the electron clockwise. The net force is

$$e(v_xH_r - v_rH_x)$$

and the torque on the electron is

$$er(v_xH_r - v_rH_x) \qquad . \quad . \quad . \quad (6\text{–}17)$$

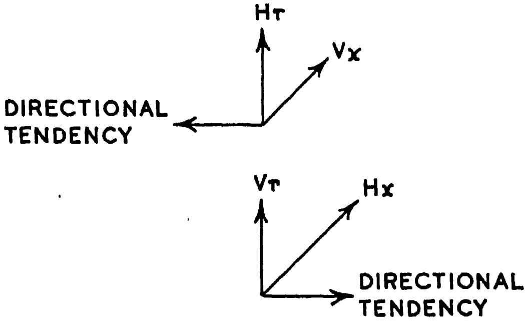


FIG. 6–15

Substituting for H_r, (6–17) becomes

$$-er\left(\frac{dx}{dt}\cdot\frac{r}{2}\frac{dH_x}{dx}+\frac{dr}{dt}H_x\right)$$

$$=-e\left(\frac{dx}{dt}\cdot\frac{r^2}{2}\frac{dH_x}{dx}+r\frac{dr}{dt}H_x\right)$$

$$=-e\left(\frac{r^2}{2}\frac{dH_x}{dt}+r\frac{dr}{dt}H_x\right)$$

$$=-e\frac{d}{dt}\left(\frac{r^2H_x}{2}\right)$$

Now, the torque on the electron is equal to $d(mr^2w)/dt$, where w is its angular velocity. Hence

$$\frac{d(mr^2w)}{dt}=-e\frac{d}{dt}\left(\frac{r^2H_x}{2}\right)$$

Integrating, there results

$$mr_2^2w_2 - mr_1^2w_1 = -e\frac{r_2^2}{2}H_{x2} + e\frac{r_1^2}{2}H_{x1}$$

where w_1, r_1, and H_{x1} are, respectively, the initial values of w, r, and H_x. If at the initial position H_x is zero, then

$$w_2 = -\frac{e}{2m}H_{x2} \quad . \quad . \quad . \quad (6\text{–}18)$$

i.e. the angular velocity of the electron at any position x_2 is proportional to the axial component of the field at that position. Thus, when the electron emerges from the field, its angular velocity is zero.

As (6–18) describes the change in the angular velocity of the electron, it is evident that the latter is rotated through a certain angle by the field. If ϕ is this angle, then

$$w = \frac{d\phi}{\ }$$

Also

$$\frac{d\phi}{dx} = \frac{d\phi}{dt}\bigg/\frac{dx}{dt}$$

$$= -\frac{e}{2m}\frac{H_x}{v_x}$$

But

$$v_x = \sqrt{\frac{2eV_0}{m}}$$

and

$$\frac{d\phi}{dx} = -\frac{1}{2\sqrt{2}}\left(\frac{e}{m}\right)^{\frac{1}{2}}\frac{H_x}{V_0^{\frac{1}{2}}}$$

Integrating

$$\phi = -\frac{1}{2\sqrt{2}V_0^{\frac{1}{2}}}\left(\frac{e}{m}\right)^{\frac{1}{2}}\int_{-a}^{+a}H_x dx$$

$$= -\frac{0{\cdot}15}{V_0^{\frac{1}{2}}}\int_{-a}^{+a}H_x dx \text{ radians} \quad . \quad . \quad . \quad (6\text{–}19)$$

where V_0 is expressed in volts. The significance of this result is that, as the angle is negative, an electron travelling in the direction of H_x will be rotated in a clockwise direction (see Fig. 6–14). Also, when an image is formed, this is rotated through an angle ϕ with respect to the object.

FOCUSING

In order to demonstrate the focusing property of the short solenoid, the radial forces acting on an electron must be considered. Due to the velocity wr, the radial force is $ewrH_x$. Also the centrifugal force is w^2rm. Substituting from (6–18) for w the resultant force is

$$-\frac{e^2r}{2m}H_x^2+\frac{e^2r}{4m}H_x^2$$

Hence $$m\frac{dv_r}{dt}=-r\left(\frac{e^2}{2m}H_x^2-\frac{e^2}{4m}H_x^2\right)$$

or $$\frac{dv_r}{dt}=-r\left(\frac{e}{2m}H_x\right)^2 \quad . \quad . \quad . \quad . \quad (6\text{–}20)$$

As the value of dv_r/dt is always negative, this result shows that electrons are always accelerated towards the axis.

FOCAL LENGTH

Multiplying (6–20) by dx, we have

$$dv_r=-\frac{r}{v_x}\left(\frac{e}{2m}H_x\right)^2dx$$

$$=\frac{rev_{x0}}{8mV_0}H_x^2\,dx$$

Integrating $$\int_{v_{r1}}^{v_{r2}}dv_r=-\frac{rev_{x0}}{8mV_0}\int_{-\infty}^{\infty}H_x^2dx$$

and $$\frac{v_{r2}}{v_{x0}}-\frac{v_{r1}}{v_{x0}}=-\frac{re}{8mV_0}\int_{-\infty}^{\infty}H_x^2dx$$

this, of course, again assuming that the transit time of the electron through the field is so short that the change in r is negligible. Referring to Fig. 6–7,

$$\frac{v_{r1}}{v_{x0}}=\tan\alpha=\frac{r}{a}$$

$$\frac{v_{r2}}{v_{x0}}=-\tan\beta=-\frac{r}{b}$$

Thus
$$-\frac{r}{b}-\frac{r}{a}=-\frac{re}{8mV_0}\int_{-\infty}^{\infty}H_x^2dx$$

and
$$\frac{1}{b}+\frac{1}{a}=\frac{0\cdot022}{V_0}\int_{-\infty}^{\infty}H_x^2dx=\frac{1}{f} \qquad . \qquad . \qquad (6\text{–}21)$$

which gives the focal length.

CALCULATION OF f

In the case of magnetic focusing, it is often easier to calculate the focal length of a magnetic lens than an electrostatic one. The field strength produced at any point on the axis of a current-carrying coil is proportional to the number of ampere-turns of the coil. Thus

$$H_x \propto N$$

Where N is the number of ampere-turns. We may, therefore, write

$$\int_{-\infty}^{\infty}H_x^2dx=(kN)^2$$

Where k is a constant depending on the coil shape. From (6–21)

$$\int_{-\infty}^{\infty}H_x^2dx=45\cdot5\,V_0\left(\frac{a+b}{ab}\right)$$

or
$$(kN)^2=45\cdot5\,V_0\left(\frac{a+b}{ab}\right)$$

and
$$N=\frac{1}{k}\sqrt{45\cdot5\,V_0\frac{l}{ab}} \qquad . \qquad . \qquad . \qquad (6\text{–}22)$$

where $l=a+b=$ the beam length.

Also
$$k=\sqrt{\int_{-\infty}^{\infty}\left(\frac{H_x}{N}\right)^2dx} \qquad . \qquad . \qquad . \qquad (6\text{–}23)$$

Now, in the case of a thin coil, where the turns are reasonably close together and the depth of winding is small compared with the radius of the coil, it is well known that*

$$H_x=\frac{2\pi r^2N}{10(r^2+x^2)^{3/2}}$$

---

[*] *Electricity and Magnetism*, p. 228, S. G. Starling (Longmans, Green).

Thus
$$\left(\frac{H_x}{N}\right)^2 = \frac{4\pi^2 r^4}{100(r^2 + x^2)^3}$$

and
$$k = \sqrt{\frac{4\pi^2}{100}} \int_{-\infty}^{\infty} \frac{r^4 dx}{(r^2 + x^2)^3}$$

To integrate this expression let $x = r \tan \theta$. We then have
$$\int_{-\infty}^{\infty} \frac{r^4 dx}{(r^2 + x^2)^3} = \frac{1}{r} \int_{-\pi/2}^{\pi/2} \frac{d\theta}{(1 + \tan^2 \theta)} = \frac{1}{r} \int_{-\pi/2}^{\pi/2} \cos^4 \theta d\theta$$

which gives $\frac{1}{r}\left[\frac{1}{4} \sin \theta \cos^3 \theta + \frac{3}{4}\left(\frac{\theta}{2} + \frac{\sin 2\theta}{4}\right)\right]_{\pi/2}^{\pi/2}$

or
$$\frac{3\pi}{8r}$$

Hence
$$k = \frac{1}{10}\sqrt{\frac{3\pi^3}{d}}$$

where d is the coil diameter. Substituting in (6–22)
$$N = \frac{10}{\sqrt{3\pi^3}} \sqrt{45{\cdot}5 V_0 \frac{ld}{ab}}$$
$$= 6{\cdot}96\sqrt{V_0 d/f}$$

MAGNIFICATION

As in geometrical optics, the ratio of image size to object size is b/a. Hence if y_i and y_0 are, respectively, the image and object diameters, the magnification is given by
$$m = \frac{y_i}{y_0} = \frac{b}{a}$$

From this result it is clear that the size of the image depends on the position of the focusing coil. For example, in the cathode-ray oscillograph a small sharply defined image of the anode aperture is desired upon a fluorescent screen. This indicates that the coil should be near the screen. However, a large distance of the coil from the anode means a large electron beam diameter at the coil. This results in the outermost electrons being focused at different positions from those with small angles of incidence, this leading to an image lacking in sharpness. This effect is, of course, analogous to spherical aberration.

In the case of an electron microscope, high magnification is desired, and hence the coil should be near the object. This is assisted by a lens of short focal length. From (6–21) it is evident that for a small value of f, H_x should be high. This may be achieved by the employment of an iron-shrouded coil with a short air-gap, as shown by Fig. 6–16. This gives a very short strong field, as may be appreciated by comparing the two field distribution curves shown. With this form of lens, focal lengths of a few millimetres have been achieved, even when the anode accelerating voltage is as high as 70 kV.

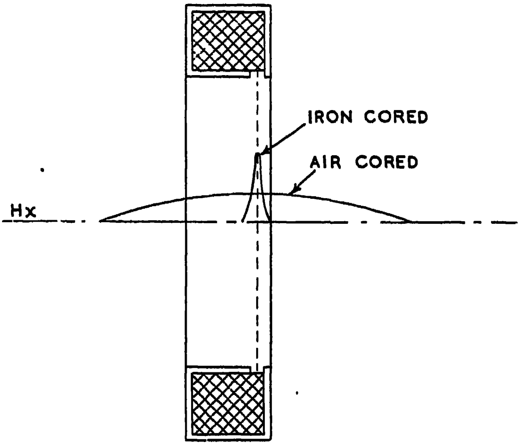


FIG. 6–16

A difference that may be noted between magnetic and electrostatic lenses is that the former cause a rotation of the image with respect to the object, the angle of rotation being given by (6–19). For cathode-ray tubes this is unimportant, because object and image are circular. Where the rotation is undesirable a magnetic lens of the type shown by Fig. 6–17 may be employed. This consists of two coils with current flowing in opposite directions. Hence the values of (6–19), for each of the fields, are equal and opposite, with the result that the net rotation is zero. The values of (6–20) are both negative, as the square of H_x is involved.

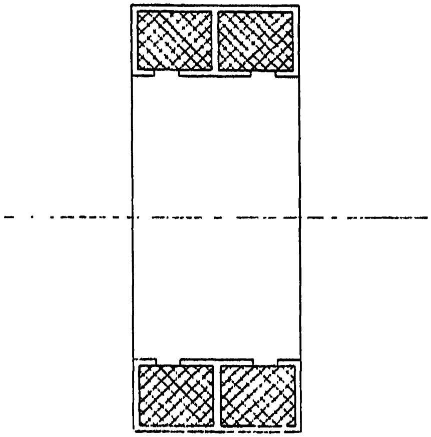


FIG. 6–17

Space-charge Effects

In discussing the differences between geometrical and electron optics on page 229 the possible effect of space-charges was mentioned.

So far this effect has been neglected, and it is now necessary to give some consideration to this matter. Due to electrons possessing similar charges, repulsion will take place between them, thus tending to cause diffusion of any beam they may form. Hence, where it is desired to form a beam of small cross-section and high density this is opposed by the mutual repulsion of the electrons forming the beam. An effect of an electron beam is to produce an electrostatic and magnetic field. The former is due to the electrons constituting the beam, and the latter to the fact that electrons in motion constitute a current.

Let n be the number of electrons in a beam passing through a plane perpendicular to the beam per second. Then if the electron velocity is v, the number of electrons per unit length of the beam is n/v. The charge per unit length is ne/v, and hence the radial electrostatic field is

$$X_x = \frac{4\pi ne}{2\pi rv} = \frac{2ne}{rv} \text{ e.s.u.}$$

Thus the radial electrostatic force on an electron at the surface of the beam is

$$\frac{2ne^2c^2}{rv} \text{ e.m.u.} \quad . \quad . \quad . \quad . \quad (6\text{–}24)$$

where c is the velocity of light.

The electrons in motion constitute a linear current, i, the magnetic field due to this being concentric with the beam and of magnitude $2i/r$, where i is expressed in absolute units. As $i = ne$, the field is given by

$$H = \frac{2i}{r} = \frac{2ne}{r} \text{ e.m.u.}$$

The radial electromagnetic force on an electron at the surface of the beam is

$$-Hev = -\frac{2ne^2v}{r} \quad . \quad . \quad . \quad (6\text{–}25)$$

the negative sign indicating that the force is directed towards the axis. The total force on a surface electron is, of course, the sum of (6–24) and (6–25), and we have

$$F = \frac{2ne^2(c^2 - v^2)}{vr}$$

It is evident from this result that the force decreases with increasing beam velocity, and thus is principally of importance where low accelerating voltages are employed when v is small compared with

c. It will also be noted that $F \propto 1/r$ and this means that it is impossible to produce point-focusing whatever optical system is adopted.

In practice, beam diffusion is generally negligible where the accelerating voltage is above 30 kV (high-voltage oscillographs, for example), but is important at voltages below 1000 V unless gas-focusing (to be shortly discussed) is employed. At these voltages the minimum diameter of the beam is limited by the beam current rather than the focusing system. In such cases it is apparent that the electromagnetic force on a surface electron is negligible compared with the electrostatic force, and hence

$$F = \frac{2ne^2c^2}{vr}$$

After the beam leaves the lens its radial velocity is v_{r2} and its acceleration, due to F, is

$$\frac{d^2r}{dt^2} = \frac{2ne^2c^2}{mvr}$$

Multiplying both sides by $2dr/dt$

$$2\frac{dr}{dt} \cdot \frac{d^2r}{dt^2} = \frac{4ne^2c^2}{mvr}\frac{dr}{dt}$$

Integrating $$\int_0^{v_{r2}} 2\frac{dr}{dt} \cdot \frac{d^2r}{dt^2} \cdot dt = \frac{4ne^2c^2}{mv}\int_{r_{min}}^{r_2} \frac{dr}{r}$$

the inferior limit of the left-hand member indicating that the radial velocity is zero where the beam radius is a minimum. Continuing

$$\left[\left(\frac{dr}{dt}\right)^2\right]_0^{v_{r2}} = \left[v_r^{\,2}\right]_0^{v_{r2}} = \frac{4ne^2c^2}{mv}\left[\log r\right]_{r_{min}}^{r_2}$$

and $$\log\frac{r_2}{r_{min}} = \frac{mvv_{r2}^{\,2}}{4ne^2c^2} = \frac{vv_{r2}^{\,2}}{4ic^2} \cdot \frac{m}{e} \quad . \quad . \quad . \quad (6\text{–}26)$$

which gives the minimum radius to which a beam may be focussed when the other quantities are known. If the radius of the beam where it leaves the lens and the distance of the image from the lens are known, then

$$\frac{v_{r2}}{v} = \frac{r_2}{b}$$

which gives v_{r2}. From (6–26) it will be noted that for a given value of r_2, r_{min} is smaller the larger the values of v and v_{r2}, and smaller the lower the value of i.

Gas Lenses

By admitting a small quantity of gas into an electron-optically focused vacuum tube it is possible to restrict considerably the diffusion of electrons from the beam. This is because the beam electrons create a positive space-charge which tends to neutralize the influence of the negative space-charge of the beam. When a sufficient positive space-charge density is produced it is possible to dispense entirely with the optical focusing system, focusing then being entirely effected by what may be termed a "gas" lens.

The early cathode-ray tubes were of the gas-focused type, focusing being effected in the following manner. The tube is first exhausted to a high degree and a small quantity of gas then introduced. The gas pressure to give optimum results depends on the gas employed, but is of the order of 10^{-3} mm. As the electrons pass up the tube, ionization by collision with gas molecules occurs. Because of the relatively large mass and low velocity of the ionized molecules, they move more slowly from the electron beam than do the secondary electrons which are produced simultaneously with them. This results in a positive charge in the path of the beam, and the resultant field due to this charge tends to cause the electrons to move towards the axis. It is evident that this phenomenon can occur in any part of the tube and hence sharp focusing may be retained when deflexion of the beam occurs under working conditions.

As previously stated, in order to obtain sharp focusing a certain gas pressure is necessary. This is because the rate of formation of ions, and hence positive space-charge density, depends upon the gas pressure. In addition to this, the formation rate depends upon the beam current and the probability of ionization of the electrons. That the space-charge density will depend upon beam current is evident because the density of the ionizing electrons is proportional to the latter. It follows from these statements that the lower the ionizing probability of the gas employed, the higher must be its pressure. Thus, while for argon the pressure is about 10^{-3} mm., for

TABLE 6-1

GAS	H_2	He	Ne	N_2	A	Kr	X
Pressure 10^{-3} mm. Hg	8 to 10	19 to 20	4·6 to 5·4	0·8 to 1·6	0·9 to 1·1	0·8 to 0·9	0·25 to 0·7

helium it is approximately 20×10^{-3} mm. Table 6–1 shows the pressure ranges of various gases to give satisfactory focusing.

In order to consider the type of beam formed when a positive space-charge is created by ionization, let ρ represent the space-charge density along the beam, ρ being assumed constant. The electron beam is considered as immersed within the positive space-charge and symmetrical with respect to the axis. The charge contained within a radius r per unit length of beam is $\pi r^2\rho$ and the flux due to this is $4\pi \,.\, \pi r^2\rho$. Hence the electrostatic field strength at r is

$$\frac{4\pi^2 r^2 \rho}{2\pi r} = 2\pi r \rho$$

The force on an electron at r is

$$-\ 2\pi r \rho e$$

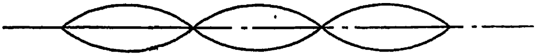


FIG. 6–18

the negative sign indicating that the electron is urged towards the axis. Equating force to mass times acceleration,

$$m\frac{d^2r}{dt^2} = -\ 2\pi r e \rho$$

or

$$m\frac{d^2r}{dt^2} + 2\pi r e \rho = 0$$

Integrating $$r = A \sin\sqrt{\frac{2\pi e\rho}{m}}\,t + B\cos\sqrt{\frac{2\pi e\rho}{m}}\,t \quad . \quad . \quad (6\text{–}27)$$

which shows that the beam radius is a sinusoidal function of the time of flight of the electron along the axis. Now, $t = x/v_0$, where v_0 is the axial velocity. Substituting

$$r = A \sin\sqrt{\frac{2\pi e\rho}{m}}\,\frac{x}{v_0} + B\cos\sqrt{\frac{2\pi e\rho}{m}}\,\frac{x}{v_0}$$

and r is a sinusoidal function of x. From this result, it is apparent that a long beam will consist of a series of nodes and antinodes as shown by Fig. 6–18. From (6–27) it will be noted that the electrons periodically cross and re-cross the beam axis with a frequency given by

$$\sqrt{2\pi e\rho/m}/\pi \text{ times per sec.}$$

the point of crossing corresponding to a node or situation at which the beam has a minimum cross-section. The distance between two nodes is

$$\frac{v_0\pi}{\sqrt{2\pi e\rho/m}}$$
$$= \sqrt{\pi V/\rho}$$

where V is the anode or accelerating voltage expressed in e.s.u.

Experimental confirmation of a nodal beam has been obtained. At very low gas pressures and accelerating voltages the beam becomes diffused after travelling a short distance. As the gas pressure is raised, focusing becomes discernible and a node-like section is formed, this being visible by virtue of the glow emitted by the excited gas molecules. For oscillograph work the gas pressure, and hence ρ, are usually arranged so that the first node is formed at the fluorescent screen.

BIBLIOGRAPHY

A Treatise on Light, R. A. Houstoun, (Longmans, Green).

"Electron Lenses," W. Wilson, *Electrical Review*, April 20th, 1945, p. 557.

"Microscopy with Light Electrons, and X-Rays," G. D. Preston, "*J. Sci. Inst.*, Dec. 1944, p. 205.

CHAPTER VII

LUMINESCENCE

UNDER suitable circumstances it is possible to produce light by means other than the age-old method of raising the temperature of a body. When light is produced by means other than heat the result is termed luminescence, whereas light produced by virtue of heat is known as temperature radiation. Hence substances raised to some temperature for the purpose of producing light are termed temperature radiators. Luminescence exists in a variety of forms and may be excited in a number of different ways. Among the earliest examples of luminescence may be mentioned the firefly and glow-worm, fungi on decaying wood, phosphorus on the sea, etc. With the exception of the last example, luminescence of this type is termed bio-luminescence. Other forms of luminescence are chemi-luminescence, photo-luminescence, and cathedo-luminescence. The first-mentioned is produced by chemical action, the second by radiation, and the third by the impact of electrons and ions. For our present purpose the two last-mentioned methods are important, for to applications of these we owe electric discharge lighting, the screens of cathode-ray tubes, television tubes, and the electron microscope and X-ray screens.

Photo-luminescence

When radiation of suitable wavelength falls on certain substances the latter may convert the former into radiation of different wavelengths. In particular, ultra-violet radiation may cause certain materials to emit radiation within the visible spectrum. This phenomenon is known as fluorescence. In some cases fluorescent materials may continue to emit visible radiation after the source of ultra-violet has been extinguished. This is known as phosphorescence, the term luminescence denoting both the phenomena of fluorescence and phosphorescence. Although, as explained below, fluorescence and phosphorescence are fundamentally the same, the first-termed is usually applied to a light emission which does not exist for a measurable time after the extinction of the excitation process. As explained above, phosphorescence denotes a light emission which exists for a definitely measurable time after excitation ceases. This time may be anything from 10^{-2} sec. to several days.

The cause of luminescence is explainable by Bohr's quantum theory of the atom; i.e. a quantum of energy may be absorbed by an atom, raising an electron from the ground state to a higher energy level, from which it ultimately returns with an emission of radiation. The method of receiving the quantum, of course, differs according to the form of luminescence. In the case of electro-luminescence (such as the glow discharge) the energy is due to the impact of an electron on a gas molecule. With photo-luminescence the energy is acquired from photons produced by the source of radiation, while with cathedo-luminescence the energy is due to the impact of an electron on some solid substance.

The time for which a fluorescent state persists is equal to the time of transition of the electron from an excited state to a lower energy level. This time is of the order of 10^{-8} sec. and is practically independent of temperature. In the case of phosphorescence the return of the electron to a lower energy level or the ground state suffers a delay. Hence considering a number of excited atoms having phosphorescent characteristics, a fraction of the electrons of these will have various times of return to the ground state, and, thus, light will be emitted for a relatively long period after the source of excitation has been extinguished. After excitation some electrons of a phosphorescent substance pass into a metastable or quasistable state of lower energy than that of the initial state to which they were raised. However, in the metastable state their energy is still higher than that of the ground state, and they ultimately return to the latter with the assistance of the thermal energy of the surrounding medium. It is during this latter process that the phenomenon of phosphorescence occurs.

The duration of the metastable state, and hence the persistence of phosphorescence, largely depends on the nature of the medium in which it occurs and the temperature. For example, in gaseous states phosphorescence does not occur. In denser media, such as liquids, short after-glows may be observed, increasing in duration and intensity with the viscosity of the liquid. With regard to the influence of temperature, phosphorescence tends to be short-lived at high temperatures, while at very low temperatures it may be "frozen in." This means that the electrons remain in the metastable state until, by a rise of temperature, the trapped energy is released as radiation.

In the majority of cases the decay of phosphorescence follows an exponential law, and in simple instances the intensity of emission may be represented by

$$I = I_0 \varepsilon^{-\alpha t} \quad . \quad . \quad . \quad . \quad (7\text{-}1)$$

where I_0 is the initial intensity, t is time, and α is a constant. The validity of this law for a single crystal of KCl(Tl) is shown by Fig. 7–1, where both the intensity and its logarithm are plotted as functions of time. If a luminescent material is not homogeneous (which is usually the case with fluorescent discharge lamps) but consists of a mixture of materials having different absorption and emission characteristics, then the total intensity of emission is generally given by a sum of exponential functions, each function being appropriate to a particular constituent of the mixture. The effect of temperature is to modify the value of α in (7–1), α increasing with a rise in temperature. Thus, the phosphorescence of calcium sulphide is observable for about thirty days at room temperature, but only for 2·8 hours at 155° C.

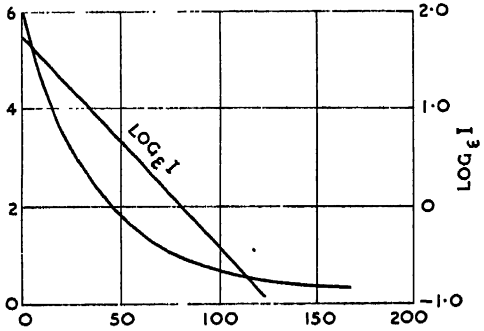


FIG. 7-1

An exponential rate of decay is not applicable to all luminescent substances, as laws of the forms $I = I_0 t^{-\alpha}$ and $I = I_0/(1 + bI_0 t)$ have been observed.

Summarizing the foregoing, it may be said that fluorescence only occurs if the medium does not influence the emission process. This means that the process occurs within the atom or molecule in a sphere of 10^{-8} cm. diameter. If the surrounding structure is involved and is modified as a result of the absorption process, time must elapse before the configuration returns to its original state. In this case phosphorescence results. It must be noted, however, that both fluorescence and phosphorescence are due to the same cause, i.e. electron transitions produced by the absorption of light quanta or photons.

STOKE'S LAW

An important empirical law was enunciated by Stokes in 1852 which states that fluorescent radiation is always of longer wavelength than that of the radiation causing it. This law may be explained with the assistance of those of the conservation of energy and the quantum theory. From the first of these laws the energy of the emitted radiation must either be equal to or less than that of the incident radiation. In liquid and solid systems it may be said that the excited molecules will transfer part of their absorbed energy to neighbouring molecules in the form of heat motion before emission occurs. If E_1 and E_2 are, respectively, the energies of the incident and emitted radiations, then

$$E_1 = hc/\lambda_1$$
$$E_2 = hc/\lambda_2$$

where λ_1 and λ_2 are the wavelengths of the incident and emitted radiations. Also

$$E_1 > E_2$$

and $$hc/\lambda_1 > hc/\lambda_2$$

or $$\lambda_2 > \lambda_1$$

which is Stoke's Law. It may be mentioned that it is the operation of this law which makes possible the conversion of ultra-violet radiation in discharge lamps into radiation within the visible spectrum. Small departures from Stoke's Law are possible at certain temperatures if heat energy is transferred from neighbouring molecules to the fluorescing molecule while it is in the excited state.

EXCITATION OF LUMINESCENCE

Substances perfectly transparent cannot be excited to luminescence by light, i.e. by visible radiation. They may, however, be excited by ultra-violet radiation should they be capable of absorbing this. Generally speaking, every photo-luminescent substance has an absorption band in the spectral region immediately adjoining the short-wave limit of the luminescent emission band. Thus, red fluorescence is excited by orange light, yellow by green, green by blue, and violet by ultra-violet.

EFFICIENCY OF LUMINESCENCE

The efficiency of luminescence is defined in three different ways: the luminous efficiency, the energy efficiency, and the quantum efficiency. The luminous efficiency is the ratio of the total light output to the total energy input and is generally expressed in lumens per watt. (See Chapter XIX.) The energy efficiency is the ratio of the energy of the luminescent light (in ergs or watt-seconds) to the energy absorbed by the luminescent material. The quantum efficiency is the ratio of the number of light quanta, or photons, emitted to the number absorbed. If η_E is the energy efficiency, then

$$\eta_E = \frac{n_2 E_2}{n_1 E_1} \quad . \quad . \quad . \quad . \quad (7\text{–}2)$$

where n_1 and n_2 are, respectively, the numbers of photons absorbed and emitted and E_1 and E_2 are the respective energies of the absorbed and emitted photons. Hence

$$\eta_E = \eta_Q \frac{E_2}{E_1} = \eta_Q \frac{\lambda_1}{\lambda_2}$$

and

$$\frac{\eta_E}{\eta_Q} = \frac{\lambda_1}{\lambda_2}$$

where η_Q is the quantum efficiency. In order that the energy efficiency shall be high, it is evident that as λ_2 is independent of λ_1, λ_1 should not be too small, i.e. the frequency of the exciting radiation should not be too far down the ultra-violet part of the spectrum. Some values of η_Q for various solid luminescent materials are given by Table 7–1.

TABLE 7–1

MATERIAL	COLOUR OF FLUORESCENCE	η_Q%
KTlCl	Ultra-violet and blue	80
ZnS(Cu) or (Ag)	Green	≒ 100
Zn_2SiO_4(Mn)	,,	70
$ZnBeSiO_4$(Mn)	Orange	25–55
$CdSiO_3$	Pink-yellow	55
$CaWO_4$	Blue	70
CdB_2O_4	Pink	66

Cathedo-luminescence

As indicated by (7–2), the energy of the primary radiation is measured by the number of photons per square centimetre and their energy, or frequency. In the case of cathedo-luminescence the exciting source consists of ions or electrons (usually the latter), and the energy is, therefore, measured by current density and the voltage through which the ions or electrons have fallen. The current density is, of course, proportional to the number of electrons passing through a square centimetre per second and thus corresponds to the number of photons in photo-luminescence. With photo-luminescence a photon can only excite a single molecule. However, an electron with sufficient energy can lose this energy stepwise and thus excite a number of molecules. Hence with cathedo-luminescence the quantum efficiency has no significance.

The minimum energy necessary for the excitation of cathodo-luminescence is much higher than might be anticipated. For luminescence within the visible spectrum the energy corresponding to limits of 7000 Å and 4000 Å is 1·6 and 2·8 electron-volts respectively. Thus, it might appear that the energy needed to excite light would be but slightly greater than the figures quoted, particularly as this is the case with photo-luminescence. Early investigations showed, however, that for the majority of solids the necessary energy is not less than about 300 electron-volts, although the energy of the emitted radiation is only a few volts.

The cause of this apparent peculiarity is of an electrical nature, rather than being an example of gross inefficiency. Where the primary electrons strike the luminescent substance secondary emission occurs. If the number of secondary electrons is less than that of primary electrons, evidently the surface of the substance must acquire a negative charge which will subsequently repel the primary electrons. Hence the latter will lose kinetic energy and arrive at the substance with far less energy than that originally possessed. Thus, the initial energy must be far larger than the few electron-volts actually necessary to cause excitation. If the secondary electron emitting properties of the substance are increased, for example, by depositing thoria on its surface, the minimum, or threshold, voltage to excite luminescence may be very much reduced.

The luminous or fluorescent intensity excited by electron impact may be written as

$$L = AN(V - V_0) \quad . \quad . \quad . \quad (7\text{–}3)$$

where N is the number of electrons received per unit area per second, V the accelerating voltage of the electrons, V_0 the threshold voltage, i.e. the lowest voltage capable of producing emission, and A is an empirical constant. The formula above is not the only one employed, other empirical formulae being

$$L = A_1J(V - V_0)^n \quad . \quad . \quad . \quad (7\text{–}4)$$

and

$$L = A_2f(J)V^2 \quad . \quad . \quad . \quad . \quad (7\text{–}5)$$

where J is the electron current density. J is, of course, proportional to N in (7–3). For small current densities, i.e. less than 5 microamps. per cm.^{2} and accelerating voltages less than 1000 volts, the appropriate formula for Willemite screens is (7–5), with $f(J)$ replaced by J. In the general formula of (7–4) the value of V_0 varies for different phosphors from 0 to several hundred volts, $f(J)$ is independent of

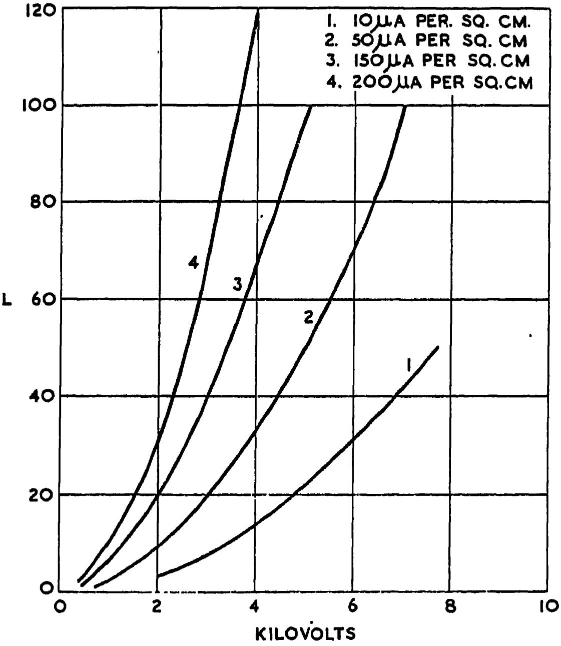


FIG. 7-2

V, and n varies from 1 to 3. Reference to Fig. 7–2 shows the variation of L with V for a typical cathode-ray tube screen and illustrates the quadratic nature of the law relating L and V. For low current densities, L is proportional to J but tends to reach a saturation value as J is continually increased. This is shown by Fig. 7–3, which concerns three different screen materials. The reason for current saturation is that there are only a limited number of centres or molecules under the influence of the electron beam and, when these are all excited, evidently further increasing J results in no further increase in L. Causes of saturation at low current densities are: the existence of relatively few luminescent centres, high excitation probabilities, and a long mean life of the excited state. Saturation does not occur with increasing voltage, because the higher the voltage the higher the energy of the impinging electrons and the greater their depth of penetration into the luminescent material. Thus, as the voltage is raised, the electrons penetrate deeper and thus excite a greater number of molecules. It is for this reason that the voltage, rather than the current density, is raised in cathode-ray tubes when greater spot intensity is desired.

Because of heating and secondary emission, the efficiency of cathedo-luminescence is very low. The energy efficiency depends on the material and voltage, but is seldom more than a few per cent. Luminous efficiencies vary from 0·25 lumen per watt for $CaWO_4$ to 3·5 lumens per watt for ZnS.

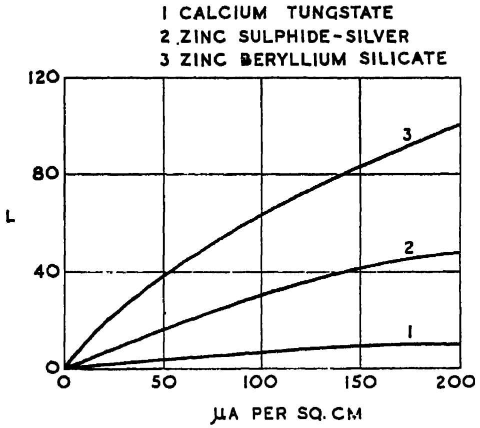


Fig. 7–3

The excitation of luminescence by X-rays is considered as an example of cathedo-luminescence rather than photo-luminescence. This is because it appears that the luminescence is excited by secondary electrons released by the X-rays rather than by the X-rays themselves. The excitation of luminescence by X-rays is chiefly employed with screens for fluoroscopy. The efficiency of such screens is of the same order as that with direct cathedo-luminescence, and in one case was found to be 3 per cent. The wavelength of the light emitted from the screen was 5250 Å and this wavelength, with an energy efficiency of 3 per cent, gives a luminous efficiency of 14·5 lumens per watt.*

PERMANENCE OF LUMINESCENCE

In some cases luminescent substances deteriorate under the action of the exciting source, the deterioration being marked by a change in colour of the substance. This effect may occur whether the source of excitation is radiation or an electron beam. The cause of deterioration is a chemical change in the luminescent substance. When the substance is in an excited state it is highly reactive and is likely to combine with any suitable material in its immediate vicinity. For example, oxidation may occur before the molecules can return to the ground state. Unless the products of combination are luminescent, the reacting molecules are lost as luminescent sources and the luminescing power of the substance declines. A further cause of loss of luminescence occurs when the exciting source is of such high energy as to dissociate the molecules.

---

* See p. 620.

In the case of cathedo-luminescence, electrons are ejected from the crystal lattice, and thus results in a partial destruction of the latter. At high current densities, of the order of 10^{-5} amp. per cm.^{2}, screens of $CaWO_4$, ZnS, and Zn_2SiO_4 may be blackened after no more than thirty minutes of constant electron bombardment of a stationary position. On cessation of the bombardment, the screen tends to recover its luminosity and colour, but nevertheless there is some deterioration. Even under normal conditions cathode-ray tube screens deteriorate and it is common to find the track swept by the spot when operating under the time-base potential only less luminous than other parts of the screen.

Luminescent Materials

A large number of substances exist which are more or less luminescent, but here we are only concerned with those of a solid nature. In the solid state no pure elements and very few simple compounds are luminescent. Solid luminescent substances may consist of either organic compounds, pure inorganic compounds, or inorganic crystals activated by inorganic impurities. For the present purpose, only the last two are of interest, particularly the latter.

In many cases pure inorganic compounds are not luminescent unless activated. In these cases luminescence depends entirely on carefully controlled small quantities of metallic impurities in the crystal lattice. Such impurities are known as phosphorogens or activators. To give some idea of the amount of activator required it may be stated that for 1 gm. of calcium sulphide about 0·00024 gm. of bismuth is necessary. Among the inorganic compounds which luminesce without activation may be mentioned zinc sulphide, ZnS, cadmium sulphide, CdS, calcium tungstate, $CaWO_4$, cadmium tungstate, $CdWO_4$, and calcium molybdate, $CaMO_4$. However, substances such as ZnS, CdS, $CaWO_4$, etc., may also be activated by a foreign impurity, in which case the colour of luminescence may be modified from that of the pure state. Thus, ZnS, which has a light-blue luminescence when unactivated, gives green luminescence when activated by copper and orange when activated by manganese.

In the case of $CaWO_4$ the luminescence does not vary with regard to colour when the activator is varied. Table 7–2 gives a list of synthetic inorganic phosphors, their various activators, colour of luminescence, etc.

TABLE 7–2

Material	Activator	Colour	Peak of Luminescence Curve, Å.	Excitation Spectrum
CaS	Mn	Orange	6000	Near U.V. to 4200
	Ni	Deep Red	7800	,, ,, ,, ,,
	Cu	Green	5100	,, ,, ,, ,,
	Bi	Blue	4750	,, ,, ,, ,,
SrS	Mn	Green-yellow	5550	,, ,, ,, ,,
	Ag	Violet	4200	,, ,, ,, ,,
	Bi	Blue-green	5200	,, ,, ,, 4400
	Ca	Blue-green	5350	,, ,, ,, ,,
ZnS	—	Light blue	4650	,, ,, ,, ,,
	Ag	Blue	4450	,, ,, ,, ,,
	Cu	Green	5200	,, ,, ,, 4300
	Mn	Orange	5850	,, ,, ,, 4500
$(Zn+Cd)S$	Mn	Orange	5900	,, ,, ,, 4600
Zn_2SiO_4	Mn	Green	5250	Below 2960
$ZnBeSiO_4$	Mn	Yellow-white	5950	,, 3000
$CaWO_4$	—	Blue	4400	,, 3000
$CdWO_4$	—	Whitish-blue	4600	,, 3000
$MgWO_4$	—	,,	4600	,, 3200

Activation and Quenching

That the number of activators for phosphors is fairly numerous is evident from Table 7–2. Also, as previously stated, the concentration of the activator is important. For example, in sulphide phosphors an increase in the activating substance increases the intensity of fluorescence relatively to the intensity of phosphorescence. Thus, to obtain phosphors with a long afterglow the activator concentration must be low. This results in weak fluorescence but strong phosphorescence. The optimum concentration of an activator for a given luminescent substance must be determined empirically, as it depends on the relative degrees of fluorescence and phosphorescence desired, whether the exciting source is photo-luminescence or cathedo-luminescence, and, in some cases, on the wavelength of the excitation. With reference to the latter consideration, in the case of $Zn_2SiO_4(Mn)$ for excitation at 3650 Å the concentration of Mn is about 4 per cent; for 2537 Å it is of the order of 2 per cent, while for 760 Å it is less than 0·5 per cent.

Where a phosphor admits different activators, the optimum concentrations of the various activators are very different. Also the concentration of the same activator employed in different materials differs. For example, the luminescence of $CaWO_4$ is quenched if the concentration of Pb exceeds 1 per cent. However, $LiWO_4$ may contain as much as 40 per cent of Pb and still be strongly luminescent. In some materials the optimum activator concentration is extremely small. As an extreme instance it may be stated that iron of the order of one part in a million will activate CaS. If the concentration is increased beyond this, the iron will quench not only the luminescence it produces but also that due to any other activator that may be present. When substances are deliberately included to have a quenching effect they are known as "killers." Phosphorescence is far more readily quenched than fluorescence and for various technical purposes killers are sometimes employed to suppress the former. Thus, where moving images are required on X-ray and cathode-ray tube screens, it is essential that there shall be a negligible afterglow. For an X-ray screen of ZnS(Cu), two parts of nickel to a million of ZnS are sufficient to destroy the phosphorescence completely while leaving the fluorescence relatively unaffected.

If the concentration of killer is too high it will tend to quench the fluorescence as well as the phosphorescence. This is shown by Table 7–3, which gives the percentage fluorescence of various phosphors for different concentrations of killers. It is assumed that the intensity of fluorescence in the absence of a killer is 100 per cent. The positive sign denotes where the impurity behaves as an activator and not as a killer.

TABLE 7–3

MATERIAL	KILLER				
	%	Fe	Ni	Cu	Mn
$Zn_2SiO_4(Mn)$	0·01	78	78	88	+
	0·1	30	16	23	+
$CaWO_4$	0·01	65	—	—	79
	0·1	37	—	—	47
$MgWO_4$	0·01	93	100	99	—
	0·1	83	100	87	—
CaS(Cu)	0·01	30	40	+	+

Electronic Applications of Luminescence

As previously stated, the most important applications of luminescence in electronics are screens and discharge lighting. For these purposes inorganic materials are exclusively used because of their superior stability. For the preparation of luminescent screens two methods are in general use. In the first method, the material is finely powdered and suspended in some liquid such as benzene or alcohol. The mixture is then evenly poured on the screen supporting plate and, when the liquid has evaporated, the sediment is left as a deposit. In the second method, the luminescent material is dusted through a sieve on to the screen support, the latter having previously been coated with a binder such as sodium silicate.

X-RAY SCREENS

The early X-ray screens consisted of barium platinocyanide, which has a strong green fluorescence. Such screens are no longer in use, partly on account of their high cost. Where stationary images only are required, suitable screen materials are the sulphide phosphors. If it is desired to show moving images, then either $CaWO_4$ or $CdWO_4$ may be employed. A disadvantage associated with these two materials is the comparatively low luminous efficiency.

The employment of tungstates instead of sulphide phosphors is, of course, due to the long afterglow of the latter. However, due to the quenching action of nickel on the afterglow of ZnS phosphors, ZnCdS activated by silver is now used almost exclusively for X-ray screens. This material gives a yellow-green fluorescence which occurs approximately at the peak of the relative visibility curve.* Hence for visual observation screens of this material are about ten times as bright as those of $CaWO_4$. A typical ZnCdS screen has the front protected by a varnish transparent to X-rays, while the support is formed of lead-glass, which transmits the fluorescence but protects the observer from the X-rays themselves.

OSCILLOGRAPH SCREENS

The material employed for oscillograph screens depends on the purpose for which the oscillograph is to be employed. For visual observation of standing figures a fluorescence is desirable which occurs somewhere near the peak of the relative visibility curve. In this case, Zn_2SiO_4 is usually employed. For photographic work a more actinic colour is desirable and in this case $CaWO_4$ is frequently used. For high-speed photography of transients a long afterglow is desirable, as this permits a relatively long exposure to be made. A long afterglow is also of value in recording slow transients such as cardiac impulses. Table 7–4 gives the approximate duration of afterglow of various materials.

---

* See page 620.

TABLE 7–4

Calcium tungstate . . .	8 μ sec.
Cadmium tungstate . . .	8 μ sec.
Willemite, Zn_2SiO_4 . . .	2–8 m. sec.
Zinc phosphate	0·25 sec.
Zinc sulphide with nickel killer .	Negligible

Television Screens

For television reproduction the image is generally required in black and white, and this requires the screen to give a white fluorescence under electron bombardment. As shown by Table 7–2, no material does this, and hence to obtain the desired result mixed phosphors must be employed. Satisfactory results may be obtained with a mixture of ZnS and ZnCdS phosphors. It is hardly necessary to state that there must be no afterglow with cathode-ray screens employed for television. This is ensured by the inclusion of a nickel "killer."

The Electron Microscope

The electron microscope is a further application of cathodo-luminescence, and in this case a magnified image is produced either on a screen or photographic plate. For visual observation ZnS activated by Cu is customary, while either ZnS(Ag) or $CaWO_4$ is employed for photography.

Fluorescent Lamps

The fluorescent lamp is an application of photo-luminescence. As explained in detail in Chapter XIX, the electric discharge within the lamp produces radiation outside the visible spectrum, and this may be converted into visible radiation if caused to act on some luminescent substance. In low-pressure mercury lamps the ultra-violet radiation principally occurs at 2537 Å, while in high-pressure lamps it occurs at 3650 Å. With neon lamps radiation occurs in the far ultra-violet at 736 and 740 Å.

The luminescent materials employed with fluorescent lamps should, as far as possible, be selected so that their maximum absorption coincides with the maximum emission of the ultra-violet radiation. For mercury-vapour lamps inorganic phosphors are employed, principally $CaWO_4$, $MgWO_4$, borates and phosphates of Cd, and Mn activated silicates of Zn and (Zn + Be + Cd). These phosphors have strong absorption and excitation bands below about 3000 Å.

Apart from advertising and purely decorative work, fluorescent lamps are generally desired to give either a daylight or sunlight effect. As with television screens, this necessitates blending of the luminescent powders. With mercury lamps, daylight effects may be produced by a mixture of 28 per cent white and 25 per cent pinkish-white $ZnBeSiO_4$ with 47 per cent $MgWO_4$. What is known as "warm white" may be produced with the same constituents, but with only 14 per cent of $MgWO_4$.

With neon discharge tubes the lines in the far ultra-violet are found to excite strongly zinc orthosilicate, Zn_2SiO_4, and calcium tungstate, the activator for the first material being Mn. The Zn_2SiO_4 gives a brilliant green fluorescence, and this combined with the normal neon discharge produces a bright yellow colour. Also the blue of calcium tungstate combined with neon gives a pink.

Further details of fluorescent lamps are given in Chapter XIX.

BIBLIOGRAPHY

"The Phosphorescence of Various Solids," J. T. Randall and M. H. F. Wilkins, *Pro. Roy. Soc.*, Vol. 184, 1945.

Fluorescence and Phosphorescence, E. Hirschlaff (Methuen).

CHAPTER VIII

HIGH-VACUUM DIODES

In dealing with the subject of thermionic emission in Chapter IV the value of the current derived was the maximum obtainable from the cathode corresponding to a particular temperature T. This current is termed the *saturation* current, but in practical arrangements is seldom attained, as the current is generally limited by what are known as space-charge conditions. Hence, in practice, the currents in valves are considerably less than saturation values and are termed space-charge-limited currents.

One of the immediate and most important applications of thermionic emission is the high-vacuum diode. Fundamentally, this is similar to the arrangement of Fig. 4-2, the cathode consisting of any one of a number of materials to be presently described. As the anode is thermionically "cold," it is evident that current can only pass in one direction through a diode. Hence it constitutes a valve and may be employed as a rectifier of alternating currents, this, incidentally, being one of the most important applications of the diode.

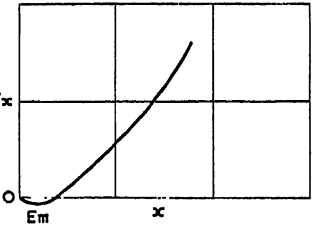


FIG. 8-1

In the absence of a sufficiently high voltage to produce saturation, the entire cathode emission does not pass to the anode. Only a fraction of the total electrons is drawn to the anode, the rest being directly returned to the cathode. This is because a negative space-charge exists in front of the cathode, the repelling effect of this on the electrons leaving the cathode having a limiting effect on the current. From the foregoing it follows that there is a negative potential gradient and potential minimum near the cathode, for only under such conditions can part of the emitted electrons be returned to the cathode. Hence the potential distribution curve between cathode and anode has the form shown by Fig. 8-1. In practice, however, the value of E_m in Fig. 8-1 is generally small compared with the cathode-anode potential, E, and hence may be neglected. Making the assumption that E_m is negligible the current-voltage relation for a non-saturated diode will now be determined.

Space-charge Equations

The case will first be considered in which the cathode and anode consist of two parallel planes situated at a distance d cm. apart, the anode being maintained at a potential $+E$ with respect to the cathode. If ρ is the negative space-charge at a distance x cm. measured perpendicularly from the cathode, then from Poisson's equation

$$\frac{d^2E_x}{dx^2} = -4\pi\rho \quad . \quad . \quad . \quad . \quad (8\text{–}1)$$

Also the current density is

$$J = -\rho v \quad . \quad . \quad . \quad . \quad (8\text{–}2)$$

where v is the velocity of the space-charge in centimetres per second. J, of course, is constant across the space between cathode and anode. As $v = \sqrt{2eE_x/m}$, from (8–1) and (8–2) we have

$$\frac{d^2E_x}{dx^2} = 4\pi J\sqrt{\frac{m}{2eE_x}}$$

Multiplying both sides by dE_x/dx,

$$\frac{dE_x}{dx}\cdot\frac{d^2E_x}{dx^2} = \frac{dE_x}{dx}\cdot 4\pi J\sqrt{\frac{m}{2eE_x}} \quad . \quad . \quad (8\text{–}3)$$

Integrating (8–3)

$$\frac{1}{2}\left(\frac{dE_x}{dx}\right)^2 = 4\pi J\sqrt{\frac{2mE_x}{e}} + C \quad . \quad . \quad (8\text{–}4)$$

where C is a constant. Now at the surface of the cathode, E_x and dE_x/dx both equal zero. Hence $C = 0$, and, from (8–4)

$$\frac{dE_x}{dx} = (8\pi J)^{\frac{1}{2}}\left(\frac{2m}{e}\right)^{\frac{1}{4}}E_x^{\frac{1}{4}}$$

Integrating $$\int_0^E E_x^{-\frac{1}{4}}\,dE_x = \int_0^d (8\pi J)^{\frac{1}{2}}\left(\frac{2m}{e}\right)^{\frac{1}{4}}dx$$

or $$J = \frac{\sqrt{2}}{9\pi}\sqrt{\frac{e}{m}}\frac{E^{\frac{3}{2}}}{d^2} \quad . \quad . \quad . \quad (8\text{–}5)$$

(8–5) gives the law relating current density and anode voltage and is of great importance in value design. It will be noted that $J \propto E^{3/2}$, and hence (8–5) is often referred to as the three-halves power law.

Alternatively, it is sometimes termed the space-charge equation. If (8–5) is expressed in practical units we have

$$J = 2{\cdot}33 \times 10^{-6} \frac{E^{3/2}}{d^2} \text{ amp. per cm.}^2 \text{ of cathode area.}$$

It will be observed that the current depends on the anode potential and electrode geometry and is independent of the cathode temperature and work function.

From the foregoing equations the values of E_x, ρ, and v may be derived in terms of x. Thus

$$E_x \propto x^{4/3}$$
$$\rho \propto x^{-2/3}$$
$$v \propto x^{2/3}$$

In practice, a more important case than the foregoing is where the emitter takes the form of a filamentary cathode surrounded by a concentric cylindrical anode. Considering Fig. 8–2, let this be represented as shown where the anode and cathode are assumed to be of unit length. If X_r is the radial field strength at r, then the difference in the total flux lines entering the leaving the annulus is

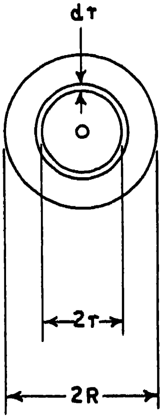


FIG. 8–2

$$d(2\pi r X_r) = 2\pi r dr \, . \, \rho \, . \, 4\pi$$
$$= d(rX_r) = 4\pi\rho r dr$$

or
$$\frac{1}{r}\frac{d(rX_r)}{dr} = 4\pi\rho$$

which is Poisson's equation in cylindrical co-ordinates.

As $X_r = dE_r/dr$, we have

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dE_r}{dr}\right) = 4\pi\rho$$

or
$$\frac{d^2E_r}{dr^2} + \frac{1}{r}\frac{dE_r}{dr} = 4\pi\rho \quad . \quad . \quad . \quad (8\text{–}6)$$

Now $J_1 = 2\pi r\rho v$

where J_1 is the current per unit length of cathode. Also

$$v = \sqrt{\frac{2eE_r}{m}}$$

and, therefore,
$$\rho = \frac{J_1}{2\pi r}\sqrt{\frac{m}{2eE_r}}$$

Substituting in (8–6)

$$\frac{d^2E_r}{dr^2} + \frac{1}{r}\frac{dE_r}{dr} = \frac{J_1}{r}\sqrt{\frac{2m}{eE_r}} \quad . \quad . \quad . \quad (8\text{–}7)$$

A particular solution of (8–7) is $E_r = Kr^a$. Then $E_r = Kr^a$, $dE_r/dr = aKr^{a-1}$, $d^2E_r/dr^2 = a(a-1)Kr^{a-2}$ where K and a are constants. Substituting in (8–7) and simplifying

$$r^{3a-2} = \frac{2mJ_1^2}{K^3a^4e}$$

Putting $a = 2/3$

$$K^3 = \frac{2mJ_1^2}{(2/3)^4e}$$

and hence $$E_r = Kr^{2/3} = \sqrt[3]{\frac{2mJ_1^2}{(2/3)^4e}} \cdot r^{2/3}$$

Thus $$E_r^{3/2} = \sqrt{\frac{2mJ_1^2}{(2/3)^4e}} \cdot r$$

and $$J_1 = \frac{2\sqrt{2}}{9}\sqrt{\frac{e}{m}}\,\frac{E^{3/2}}{R} \quad . \quad . \quad . \quad . \quad (8\text{–}8)$$

It will be noted that both in this case and in that for parallel planes the three-halves power law is obeyed. It is found that this law is approximately true no matter what the shapes of emitter and anode may be.

The solution of (8–7) was obtained by assuming that the relation between E_r and r is of the same form as that found for parallel planes. A general solution of this equation is

$$J_1 = \frac{2\sqrt{2}}{9}\sqrt{\frac{e}{m}}\,\frac{E^{3/2}}{\beta^2R}*$$

where β is a function of r_1/R. However, for most cases found in practice $\beta \fallingdotseq 1$. Fig. 8–3 shows β^2 plotted against r_1/R.†

If (8–8) is expressed in practical units there results

$$J = 14{\cdot}65 \times 10^{-6}\,\frac{E^{3/2}}{R} \text{ amp. per cm. length of cathode.}$$

Effect of Cathode-potential Drop

The foregoing space-charge equations have been derived on the assumption that the cathode is an equipotential surface. This assumption, while true for indirectly-heated cathodes, is incorrect when the cathode is directly heated for, in the latter case, the potential continuously changes from one end of the cathode to the other. Thus, if a battery is employed for heating the cathode, the potential difference between the ends of the latter is, of course, equal to the battery voltage. An effect of the cathode potential drop is a modification of the three-halves power law and this modification must now be considered.

---

* This is sometimes termed Langmuir's Equation.

† r_1 is the filament radius.

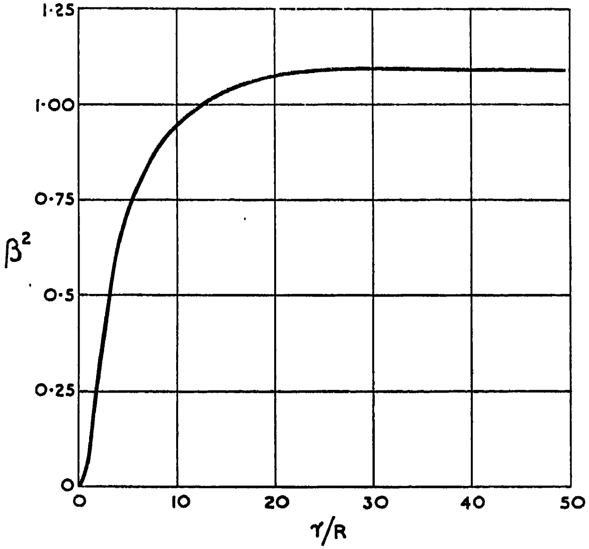


Fig. 8–3

Consider a filamentary cathode, concentric with a cylindrical anode, and a steady potential difference E_c maintained between the cathode extremities, as shown by Fig. 8–4. If the negative end of the cathode is taken to be at earth potential, then the potential at a point x along the cathode is E_{cx}. Assuming the space-charge equation to apply to every elementary length of the cathode, the contribution to the anode current from the element dx is

$$dI = K(E - E_{cx})^{3/2}\,dx \quad . \quad . \quad . \quad (8\text{-}9)$$

where K is constant depending on the geometry of the electrode system. As E_{cx} is proportional to x

$$\begin{matrix} E_{cx} & x \\ E_c & l \end{matrix}$$

and substituting in (8–9)

$$dI = K\left(E - \frac{x}{l}E_c\right)^{3/2} dx$$

Integrating
$$\int dI = \int_0^l K\left(E - \frac{x}{l}E_c\right)^{3/2} dx$$
$$= -\frac{2}{5}K\frac{l}{E_c}\left[\left(E - \frac{x}{l}E_c\right)^{5/2}\right]_0^l$$
$$= \frac{2}{5}K\frac{l}{E_c}\left[E^{5/2} - (E - E_c)^{5/2}\right]$$

or
$$J_1 = \frac{2}{5}\frac{K}{E_c}E^{5/2}\left[1 - \left(1 - \frac{E_c}{E}\right)^{5/2}\right] \quad . \quad . \quad (8\text{–}10)$$

Now E_c is usually small compared with E. Hence, we may expand the term in parentheses in (8–10) by the binomial theorem, retaining only the first three terms, with the result

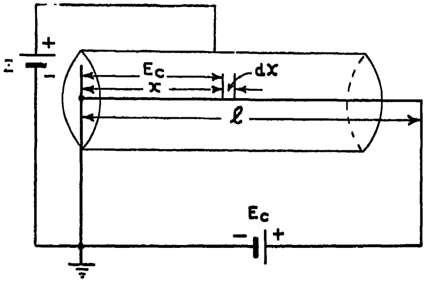


FIG. 8-4

$$J_1 = KE^{3/2}\left(1 - \frac{3}{4}\frac{E_c}{E}\right) \quad . \quad . \quad . \quad (8\text{–}11)$$

which, of course, reduces to (8–5) or (8–8) if E_c is negligible. Where E_c is comparable with E, the expression (8–10) must be employed.

In Fig. 8–4, the anode current enters the cathode at its negative extremity. Should it be returned to the positive end then, with the same approximation as in (8–11),

$$J = KE^{3/2}\left(1 + \frac{3}{4}\frac{E_c}{E}\right) \quad . \quad . \quad . \quad (8\text{–}12)$$

It is evident from (8–11) and (8–12) that, for a given value of E, the magnitude of the anode current will depend on whether it is returned to the negative or positive end of the cathode, being greater in the latter case. In the first case the current is less than that obtainable from an equipotential cathode, and in the second case, greater.

In order to simulate the effect of an equipotential cathode, the arrangement of Fig. 8–5 may be employed where two equal resistances *R* are shunted across the cathode, as shown. Under these circumstances, providing the cathode heating current is large compared with the anode current, the centre of the cathode and the junction of the resistances will be at the same potential, i.e. $E_c/2$. Hence the emission from the left-hand half of the cathode occurs at a lower potential than that from the right-hand. Taking the origin as the centre of the cathode in Fig. 8–5, the current due to the left-hand side is

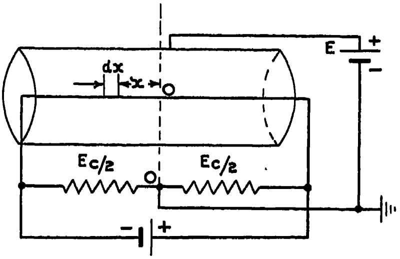


FIG. 8–5

$$\int dI_l = \int_0^{l/2} K\left(E - \frac{x}{l}E_c\right) dx$$

$$= -\frac{2}{5}K\frac{l}{E_c}\left[\left(E - \frac{x}{l}E_c\right)^{5/2}\right]_0^{l/2}$$

and that due to the right-hand side

$$\frac{2}{5}K\frac{l}{E_c}\left[\left(E + \frac{x}{l}E_c\right)^{5/2}\right]_0^{l/2}$$

Hence $$J_1 = \frac{2}{5}\frac{K}{E_c}\left[\left(E + \frac{E_c}{2}\right)^{5/2} - \left(E - \frac{E_c}{2}\right)^{5/2}\right]$$

Expanding by the binomial theorem, all the odd terms disappear, and we have

$$J_1 = KE^{3/2}\left[1 + \frac{1}{32}\left(\frac{E_c}{E}\right)^2 + \frac{3}{10240}\left(\frac{E_c}{E}\right)^4 + \ldots\right]$$

From this result it is evident that with the arrangement of Fig.

8–5, the anode current is slightly higher than with an equipotential cathode. However, because of the rapid convergency of the series and the fact that, in practice, E_c is usually small compared with E, the arrangement of Fig. 8–5 may be considered to be almost exactly equivalent to an equipotential cathode.

A.C. Directly-heated Cathodes

In the majority of instances cathodes are heated by alternating current and in such instances the anode current can be returned via a centre tap on the cathode-heating transformer winding. Whether the anode current is returned in this manner or to one end of the cathode, the foregoing equations for the d.c. examples are valid, provided for E_c we write E_{cm} sin pt, where E_{cm} is the maximum value of the cathode voltage. Thus, for the centre-tapped case

$$J_1 = KE^{3/2}\left[1 + \frac{1}{32}\left(\frac{E_{cm}}{E}\right)^2 \sin^2 pt + \ldots\right] \quad . \quad . \quad . \quad (8\text{–}13)$$

while for a connexion to the cathode extremity

$$J_1 = KE^{3/2}\left[1 - \frac{3}{4}\frac{E_{cm}}{E}\sin pt\right] \text{ or } KE^{3/2}\left[1 + \frac{3}{4}\frac{E_{cm}}{E}\sin pt\right] \quad (8\text{–}14)$$

Either of the right-hand members in (8–14) is applicable, for changing the connexion from one end of the cathode to the other merely causes a phase reversal.

Considering (8–13) and (8–14), it will be noted that in both cases an alternating component is superposed on the anode current, the value of this component being much smaller when the centre-tapped arrangement is employed. As $\sin^2 pt = \frac{1}{2}(1 - \cos 2pt)$ the frequency of the superposed component in (8–13) is twice that of the fundamental.

In Chapter X it is shown that a triode is equivalent to a diode provided that E is replaced by $E + \mu E_g$, where μ is the amplification factor and E_g the cathode-grid potential. Hence the foregoing equations are applicable to triodes if E is replaced by $E + \mu E_g$. As μE_g may be of the same order as E, it will be seen that substitution of $E + \mu E_g$ in (8–14) indicates the possibility of a relatively large a.c. component in J_1 with the likelihood of an audible hum in an audio-frequency amplifier operated with valves having a.c. directly-heated cathodes. Hence in these circumstances a centre-tapped heater winding should be employed, when (8–13) will be applicable.

Modifying Factors of the Space-charge Equation

If a gradually increasing anode voltage is applied to a diode, it appears that the anode current should increase in accordance with the three-halves power law until limited by saturation. Thus when the emission reaches the saturation value corresponding to the particular cathode temperature, the curve should abruptly cease to rise and then lie parallel to the axis in the manner shown by Fig. 8–6. In practice, the discontinuity in the curve of Fig. 8–6 does not exist for several reasons. For example, the cathode temperature is not constant, being higher at the cathode centre than at the ends. Thus saturation occurs earlier for the ends of the cathode than for the centre, this causing the curve to have the appearance shown by Fig. 8–7. The variation in temperature along the cathode, of course, causes the resistance of the latter to be non-uniform and hence E_{cx} does not in fact vary strictly as x/l. For this reason the equations (8–11) to (8–14) are only approximately correct.

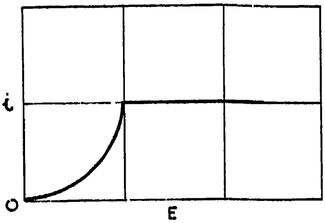


Fig. 8–6

A further factor influencing the cathode temperature is the emission occurring therefrom. As shown on page 177, the power loss due to electron emission is $J(\phi + 2\phi_r)$ per cm.^{2} of cathode surface and this has a cooling effect on the cathode. Thus, as the anode current is increased, the cathode temperature tends to fall, causing saturation to occur earlier than if the cathode temperature remained constant.

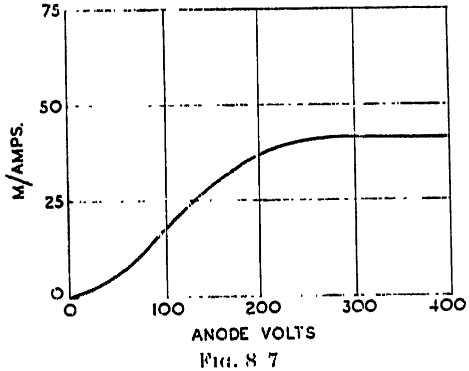


Fig. 8–7

Vacuum Valve Design and Construction

In order that vacuum diodes shall be suitable for the various purposes for which they are employed, a high degree of vacuum must be attained. This is essential to limit the production of gaseous ions which would be detrimental to the diode's characteristics and life. The number of gas molecules per cubic centimetre at normal temperature and density is $2{\cdot}7 \times 10^{19}$, and at any other density ρ the number is given by

$$2{\cdot}7 \times 10^{19}\rho/\rho_0 \text{ per cm.}^3$$

where ρ_0 is the gas density under normal conditions. During the passage of the electrons across the inter-electrode space, these may collide with residual gas molecules, thereby producing ions by collision. As the diameter of a gas molecule is of the order of 10^{-8} cm., the area which it presents to an oncoming electron is of the order of 10^{-16} cm.^{2} If each collision results in a positive ion, then, if d is the distance between cathode and anode, the number of ions per electron is

$$2{\cdot}7 \times 10^{19}\frac{\rho}{\rho_0} \times 10^{-16} \times d$$

$$= 2{\cdot}7 \times 10^{3}\frac{\rho}{\rho_0} d$$

Now, the velocity of the electrons is greater than that of the ions in the ratio of $\sqrt{1839M} : 1$, where M is the molecular weight of the gas. Hence the ratio of positive ion density to electron density may be taken as

$$= 2{\cdot}7 \times 10^{3}\frac{\rho}{\rho_0} d\sqrt{1839M}$$

and taking an upper limit for this of 10^{-2}

$$\frac{\rho}{\rho_0} d\sqrt{M} = 10^{-7}$$

For the case of air $M = 40$, and if $d = 1$ cm., then $\rho = 10^{-5}$ mm. Hg. In practice this is the order of vacuum attained. Even at this pressure the number of molecules per cubic centimetre is about 4×10^{11}, the mean free path being about 10^3 cm.

Cathode Materials

The subject of cathode materials has been lightly touched upon in Chapter IV, and it is now necessary to amplify what has already been said on this subject. Although a great number of materials emit electrons when heated, few are suitable as practical cathodes, and, in practice, three materials are principally employed. These are tungsten, thoriated tungsten, and certain oxides. As stated in Chapter IV, the efficiency of a cathode is expressed by

$$\frac{\text{Thermionic Current Obtainable}}{\text{Cathode Heating Power}}$$

and this ratio is usually given in milliamperes per watt. Evidently it is desirable that this ratio shall be as high as possible consistent with other requirements, such as long cathode life, etc.

TUNGSTEN CATHODES

In addition to high-emission efficiency other desirable characteristics for a cathode are long life, a maximum emission which is considerably above the normal working emission, and immunity from damage by positive ion bombardment. Among the earliest materials employed for thermionic cathodes was tungsten, a material which is still almost exclusively employed in high-voltage tubes. The reason for this employment is that tungsten more readily withstands positive ion bombardment than do the other two materials mentioned above. In high-vacuum valves the volt drop is relatively high because of the presence of the negative space-charge. Because of this high drop, the few positive ions which are formed acquire high velocities and thus may bombard the cathode with considerable energy. The coatings of the coated-cathodes, to be shortly described, would be rapidly destroyed by this bombardment and for this reason it is necessary to employ tungsten.

The melting-point of tungsten being 3650° K. enables it to be operated at a relatively high temperature, about 2400° K. By working at such temperatures a higher total emission is, of course, realizable than if a lower temperature were employed. However, the emission efficiency of tungsten is low, about 2 mA per watt at 2400° K., the total emission per watt at this temperature being about 120 mA per cm.^{2} These figures relate to cathodes of filamentary form of 0·1 mm. diameter approximately. For cathodes of greater diameter better figures are obtainable because larger diameters may be run at higher temperatures.

THORIATED-TUNGSTEN CATHODES

From Table 4–1, Chapter IV, the values of A and ϕ for tungsten are 60·2 and 4·52, respectively. Compared with various other metals in the table it is evident that the work function of tungsten is high. Hence in order to obtain a high-emission efficiency, it is necessary to employ a cathode material of low-work function. If a material of lower work function than tungsten were substituted for the latter in a vacuum diode, all other conditions remaining the same, there would be no increase in anode current for, as we saw on page 275, this is governed by the space-charge equation. However, the same current could be obtained with a lower expenditure of cathode-heating power, and, moreover, the *total* emission would be greater.

The low-work function metals are in general unsuitable for direct employment as cathodes, for they possess low melting-points and high vapour pressures. In practice these difficulties are overcome by depositing a layer of low-work function material on some base metal, such as tungsten, nickel, etc., in which case the material is absorbed by the base metal. Thus, the absorbed material may be held within the base metal at a temperature at which it would normally rapidly evaporate.

An example of the foregoing process is thoriated-tungsten. To prepare cathodes of this material between 1 and 2 per cent of thoria, i.e. thorium oxide (ThO_2), is added to the tungsten powder before it is sintered and drawn into wire. After the filament has been formed and assembled into the valve it must be activated in order to obtain high emission. The activation process is effected in the following manner: The filament is raised to a temperature of 2800° K. for several minutes, during which time some of the thorium oxide is reduced to metallic thorium. At this temperature the thorium diffuses to the surface of the filament, where it rapidly tends to evaporate. The temperature is now reduced to 2100° K., at which value the thorium still diffuses to the filament surface, but with a reduced rate of evaporation. Because of the latter, the thorium tends to accumulate at the surface, where it forms an absorbed layer. The activation process is now complete, and the filament so formed has an emission approximately a thousand times as great as a pure tungsten filament operating at the same temperature. The normal operating temperature of a thoriated-tungsten cathode is about 1900° K., at which temperature the rates of diffusion and evaporation are such that the absorbed layer remains fairly stable.

With use and the progress of time the layer of thorium atoms slowly evaporates and this is accompanied by a loss of emission. In order to reduce the rate at which this occurs and thus increase the longevity of the valve, thoriated-tungsten cathodes are usually carbonized. In this process the cathode is heated to a temperature of about 1800° K. in the presence of naphthalene or benzol, and, as a result, some of the carbon diffuses into the tungsten. The effect of this is to reduce the rate of evaporation of thorium to about one-sixth of that occurring with a non-carbonized cathode.

The principal effect of the layer of thorium on the surface of a thoriated-tungsten cathode is a lowering of the potential barrier and work function. The value of the latter is lower than for either pure tungsten or thorium, although it is difficult to give a precise value for this quantity as this depends on the fraction of the surface area covered by thorium. The same difficulty exists regarding the value of A, for which values ranging from 3 to 59 amp./(cm.^{2})(deg.^{2} K.) have been given by various experimenters. At a temperature of 1900° K. the emission efficiency is about 40 mA per watt and the total emission about 1·2 amp. per cm.^{2} Thus, the emission efficiency is some twenty times greater than that of pure tungsten, while the total emission is about ten times as great.

OXIDE-COATED CATHODES

An even higher emission efficiency than that of thoriated-tungsten is obtained from oxide-coated cathodes. In order to form these cathodes certain oxides, such as those of barium and strontium, are deposited on a metal core. This core may consist of an alloy of platinum and nickel, nickel and cobalt, or nickel, cobalt, iron, and titanium. In the case of indirectly-heated valves the metallic core is in the form of a sleeve. The oxide coating is applied to the core either by dipping or spraying, an organic binder being mixed with the oxide to make it adhere to the core. Thus, one method is to dip the core in a mixture of barium and strontium carbonates and resin, the core being subsequently momentarily heated to 1000° K. to burn off the binder. After an adherent coating has been applied, the cathode is heated to about 1300° K. for several hours, during which time the carbonates are reduced to oxides.

Following the formation of the oxide coating, the cathode must be activated before appreciable emission can be obtained. This process consists of maintaining the cathode temperature at between 1000° and 1500° K., with a positive anode potential to collect the emission. During the process the emission increases and oxygen is evolved, the latter being removed by the pump employed for evacuating the glass envelope.

The action of the oxide-coated cathode is somewhat obscure. It is thought that a thin layer of barium is formed, but it is not certain whether this formation occurs on the core or on the surface of the oxide. The view is generally held that it is the barium which is responsible for relative high emission, and not the oxides, this idea being supported by the evolution of oxygen during the activation process. However, whatever may be the exact action of such cathodes, the composite structure results in a low potential barrier and work function. For similar reasons, as with thoriated-tungsten, it is difficult to give precise values for A and ϕ with oxide-coated cathodes. Despite this, however, figures frequently quoted are $A = 0{\cdot}01$ amp./(cm.^{2}) (deg.^{2} K.) and $\phi = 1{\cdot}0$ volt, the value for A being derived from the fact that, at 1000° K., the emission density is usually of the order of 1 mA per mm.^{2} The emission efficiency is of the order of 100 mA per watt, the normal operating temperature being 1000° K.

As with thoriated-tungsten cathodes, oxide-coated cathodes become de-activated with use. Similarly, de-activation occurs with positive ion bombardment, and for this reason oxide-cathodes are only employed in valves having relatively low anode voltages. Because the oxides tend to evolve gas, valves with such cathodes possess a relatively large number of positive ions and are not used where the anode potential exceeds 1000 volts.

INDIRECTLY-HEATED CATHODES

The various cathodes so far described are of the directly-heated type, i.e. the heating current passes directly through the cathode. As we have already seen, this results in a non-equipotential cathode surface which, in certain applications, is detrimental. To avoid this, indirectly-heated cathodes are now widely employed. With this form of cathode the active coating is deposited on a nickel or nickel-alloy sleeve, the heater wire being contained in an insulator which is enclosed by the sleeve. Some typical arrangements are shown by Fig. 8–8. The entire structure is of such mass that temperature fluctuations are negligible at normal supply frequencies. The heater wire is usually formed of tungsten and being insulated from the metal sleeve results in the latter having an equipotential surface.

The insulator is usually formed of beryllium or aluminium oxide, such materials having relatively high thermal conductivity. For example, the thermal conductivity of the latter material is five times as high as that of porcelain and its thermal expansion is relatively close to that of nickel. The heater wire is coated with the oxide as a fluid mixture, an organic binder being present in the latter. The coated heater is then dried and is afterwards heated to about 1300° K. in a vacuum. Following this the coated heater is inserted and fixed within the metal sleeve.

A further advantage of the indirectly-heated cathode is that, because of its rigid structure, smaller clearances may be employed, this leading to a higher mutual conductance in a triode than when directly-heated valves are used.* Thus, comparing similar valves, a directly-heated general-purpose triode might have a mutual conductance of 1·5 mA per volt, while its equivalent indirectly-heated type would have a figure double this value. A disadvantage of the indirectly-heated cathode is its low-emission efficiency. Thus, the cathode consumption of the first of the two valves mentioned is only 0·2 watt, while the second consumes 4 watts. This, however, is of no great consequence unless the cathodes are to be supplied by batteries.

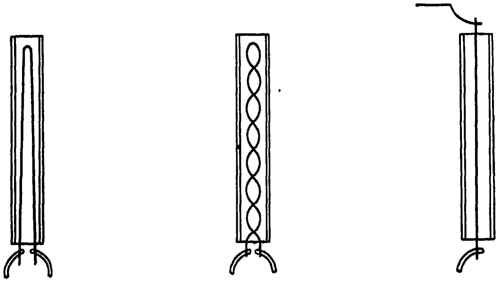


FIG. 8–8

As the cathode is insulated from the heater, it follows that three leads will be brought out from the cathode structure. The heater is not, of course, regarded as an electrode, but, because of the insulating coating, several heaters may be supplied from a common source, although their respective cathodes may be at different potentials. However, it must be borne in mind that the insulating layer is thin and that the heater and cathode should not be subjected to a potential difference of more than 500 volts under normal operating conditions.

HEAT-SHIELDED CATHODES

Another extremely important form of cathode is the heat-shielded type. However, as this type cannot be employed in vacuum valves, a discussion of its characteristics is reserved for Chapter IX, page 300.

---

* See page 317.

The Anode

During the passage of the anode current in a diode heat, is generated at the anode surface. The rate of production of heat is equal to ie, where i and e are, respectively, the instantaneous values of the anode current and voltage. This heat is, of course, due to the loss of kinetic energy by the electrons when they strike the anode. The resultant temperature of the anode is not only due to the cause just mentioned, but also due to the fact that some of the heat radiated from the cathode is intercepted by the anode. Thus, the latter must radiate energy at a rate equal to

$$P = ie + \beta E_c I_c$$

where E_c and I_c are, respectively, the cathode voltage and current and β is a number less than unity. The final temperature of the anode will, of course, be such that it radiates energy at the same rate as that at which it receives energy. As the anode is situated in a vacuum, the final temperature will be governed by the Stefan-Boltzmann Law,* i.e.

$$ie + \beta E_c I_c = \sigma\gamma T^4$$

From this equation it is evident that the final temperature will depend on the value of γ, the emissivity coefficient, this quantity in turn depending on the nature of the anode material. Materials suitable for anodes must be such as to withstand the heat generated both in the process of manufacture and in use. Furthermore, they must not emit gases which will reduce the degree of vacuum of the valve.

The anode materials in general use are nickel, iron, tantalum, molybdenum, tungsten, and graphite. The first two are employed for valves of small capacity and the latter four for large power valves. Frequently nickel anodes are roughened and blackened with carbon in order to increase their emissivity. In high-power valves the anode is sometimes operated at a red heat. Where a diode is employed as a rectifier, the potential of the anode periodically reverses. Hence it is important that the anode temperature should not be such as to permit thermionic emission to occur, for should this take place the rectifying properties of the valve would be impaired. The maximum permissible rates of energy dissipation of nickel, molybdenum, and tungsten have been given as 3, 5, and 8 watts per cm.^{2}, respectively.

As the power rating of a valve increases, so also do the losses, and, in consequence, it becomes increasingly difficult to dissipate the latter without making the valve of excessive size. To overcome this difficulty, valves having external anodes have been developed. Such valves have walls formed partly of glass and partly of metal, this construction necessitating a metal-to-glass seal. The anode is cooled either by water or an air-blast.

---

* See page 177.

Electrode Leads

Apart from the necessity of employing leads with a thermal expansion similar to that of glass, no special difficulty exists regarding the electrode connexions of low-voltage valves. However, with high-voltage valves the possibility of a flash-over exists, either within the glass pinch or outside, if the anode and cathode leads are both brought out from the same end of the valve. A further point is that with a high-potential gradient conduction of an electrolytic nature may occur within the glass, particularly when this is hot. Where the voltage of a valve exceeds about 1000 volts, it is customary to bring the cathode and anode leads out from opposite ends of the valve.* The high voltage, of course, does not exist between the electrode leads while conduction through the valve is occurring. However, with valves employed for rectification, it is shown in Chapter XIII that during the reverse half-cycle a voltage as high as twice the peak voltage of the a.c. supply may be impressed on the valve electrodes. As high-vacuum diodes are employed in X-ray work to provide d.c. voltages up to 250,000 volts, it will be appreciated that a diode may have to withstand an inverse peak voltage of the order of half a million volts. In extreme circumstances such valves are immersed in oil to increase their resistance to flash-over.

---

* See Fig. 13-13.

CHAPTER IX

GAS-FILLED DIODES AND TRIODES

THE introduction of gas into a thermionic valve may have a profoundly modifying effect on its characteristics. It can be stated at once that the principal effect is a reduction in potential for a given current. In the absence of gas or vapour in any appreciable quantity, the voltage-current characteristic is, of course, similar to that shown by Fig. 8–7, and is governed by Langmuir's and Richardson's equations. If, now, a small quantity of gas or vapour is introduced (but much smaller than is customarily employed), the characteristic will be altered from Curve 1 to Curve 2 in Fig. 9–1, Curve 1 being merely a reproduction of that in Fig. 8–7. Increasing the gas pressure still further will produce the characteristic 3, while, if the pressure is continually increased, the voltage may even fall with increasing current.

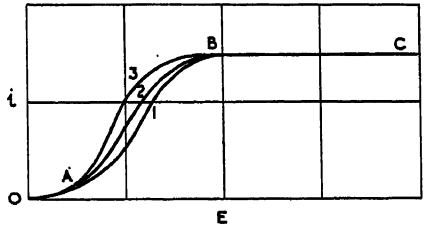


FIG. 9–1

From Fig. 9–1 it is evident that the effect of the gas is to modify that part of the characteristic lying between *A* and *B*. The portion *BC* is scarcely affected, for the current in this region is mainly limited by the maximum emission of the cathode, which is the same whether gas is present or not. The region *OA*, where the voltage is less than the ionizing potential of the gas, is also but slightly affected. The modification of *AB* is, of course, due to ionization occurring within the gas. As soon as the cathode electrons have acquired sufficient velocity and energy they ionize the gas, producing positive ions and further electrons in equal numbers. The effect of this is twofold: first, because of their motion, the ions and electrons increase the current; and second, because of their charge, the ions tend to neutralize the negative space-charge. The first effect is of considerably less importance than the second, as the number of positive ions produced is only a small percentage of the number of electrons, which is shown by the inappreciable increase in saturation current when the gas is present.

Although the effect of the gas is relatively unimportant in the region *OA*, nevertheless it may have a modifying effect on this part of the characteristic. Below the ionizing potential, the gas may retard the motion of the electrons, thus tending to reduce the current for a given voltage. Where a gas has a metastable state, this and the initial velocity of the electrons may cause the current in the region *OA* to be greater than for the case of a vacuum. However, in general, it may be said that the effect of the gas becomes appreciable only at voltages of the order of the ionizing potential.

In a similar manner to the glow discharge, described in Chapter II, the positive ions tend to concentrate the tube voltage drop into a cathode fall close to the cathode. With a thermionic cathode, however, there is a copious supply of electrons and, unlike the glow discharge, the minimum cathode fall is not limited by Townsend conditions. For the maintenance of the discharge all that is necessary is that the electrons shall fall through a sufficient potential difference to acquire the energy necessary to ionize the gas. Hence, gas-filled thermionic diodes are characterized by a potential drop of about 5 to 20 volts and this is roughly independent of current and tube geometry.

Conduction Conditions in Gas-filled Thermionic Diodes

An attempt will now be made to present a somewhat more quantitative representation of the operating conditions during conduction in the gas-filled diode. As the potential difference across the electrodes is gradually increased from zero, the current is initially carried entirely by electrons and the conditions are governed by the laws of a vacuum diode until the ionization potential of the gas is reached. Beyond this, the current increases more rapidly until its value is such that a voltage maximum is reached. From this point the current increases with great rapidity to a value limited only by the external circuit resistance or the total emission of the cathode. The tube is said to *breakdown*, *strike*, or *fire*, and its appearance is characterized by a visible glow the colour of which is appropriate to the particular gas or vapour filling. The volume of the glow constitutes a plasma similar in character to that of the glow discharge described on page 90. Between the plasma and the electrodes and walls of the tube are sheaths.

The distribution of potential over the cathode-anode space after striking has occurred is shown by Fig. 9–2. The region *OA* is known as the positive-ion or cathode sheath and the region *AB* the plasma. An explanation of this distribution may be obtained as follows. Consider the relatively simple case where anode and cathode consist of two parallel planes situated d cm. apart. Provided the voltage of the anode relative to the cathode is slightly above the ionization potential of the gas, positive ions will form at the anode, thence moving towards the cathode. If each electron produces α ions then, provided there is no recombination, the electron current density is J_- and the positive ion current density $J_+ = \alpha J_-$. If E_x is the potential at a point x cm. from the cathode, then

$$\tfrac{1}{2}m_1 v_1^2 = eE_x \quad . \quad . \quad . \quad . \quad (9\text{–}1)$$

$$\tfrac{1}{2}m_2 v_2^2 = e(E - E_x) \quad . \quad . \quad . \quad (9\text{–}2)$$

where m_1, v_1 and m_2, v_2 are, respectively, the masses and velocities

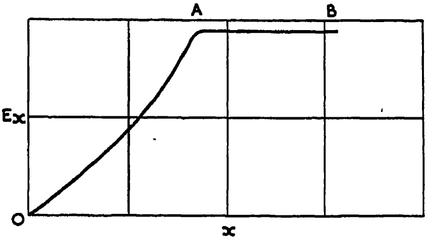


FIG. 9–2

of the electrons and ions, E is the anode potential, and e the electronic charge. Now

$$J_- = \rho_1 v_1 \quad . \quad . \quad . \quad . \quad (9\text{–}3)$$

$$J_+ = \rho_2 v_2 \quad . \quad . \quad . \quad . \quad (9\text{–}4)$$

where ρ_1 and ρ_2 are, respectively, the electronic and ionic densities. Also we have

$$\frac{d^2E_x}{dx^2} = 4\pi(\rho_1 - \rho_2) \quad . \quad . \quad . \quad (9\text{–}5)$$

From (9–1), (9–2), (9–3), and (9–4), (9–5) may be written

$$\frac{d^2E_x}{dx^2} = 4\pi\left[J_-\sqrt{\frac{m_1}{2e}}\,E_x^{-\frac{1}{2}} - J_+\sqrt{\frac{m_2}{2e}}\,(E - E_x)^{-\frac{1}{2}}\right]$$

$$= 4\pi J_-\sqrt{\frac{m_1}{2e}}\,[(E_x^{-\frac{1}{2}} - K(E - E_x)^{-\frac{1}{2}})] \quad . \quad . \quad (9\text{–}6)$$

where $K = \dfrac{J_+}{J_-}\sqrt{\dfrac{m_2}{m_1}}$

Multiplying both sides of (9–6) by dE_x/dx and integrating,

$$\frac{1}{2}\left(\frac{dE_x}{dx}\right)^2 = 4\pi J_-\sqrt{\frac{2m_1}{e}}\,[E_x^{\frac{1}{2}} + K(E - E_x)^{\frac{1}{2}}] + c$$

where c is a constant.

At the cathode surface x, E_x, and $dE_x/dx = 0$, and therefore

$$c = 4\pi J_- \sqrt{\frac{2m_1}{e}}\, K E^{\frac{1}{2}}$$

and
$$\left(\frac{dE_x}{dx}\right)^2 = 8\pi J_- \sqrt{\frac{2m_1}{e}}\, [E_x^{\frac{1}{2}} + K(E - E_x)^{\frac{1}{2}} - KE^{\frac{1}{2}}] \qquad (9\text{–}7)$$

At the anode $E_x = E$ and in these circumstances

$$\left(\frac{dE}{dx}\right)^2 = 8\pi J_- \sqrt{\frac{2m_1}{e}}\, E^{\frac{1}{2}}(1 - K)$$

As dE/dx must be real and positive, it is evident that $K \leq 1$.

Now, if $K = 0$, (9–7) reduces to (8–4) and the current is given by (8–5). This is, of course, the condition until striking occurs.

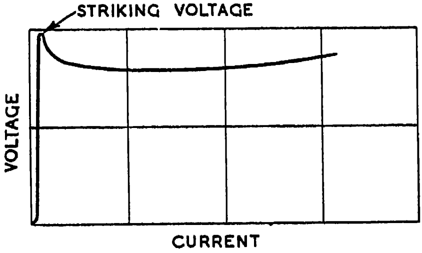


FIG. 9–3

After striking has occurred a plasma is formed and within this $\rho_1 = \rho_2$ and $K = 1$ *within* the plasma. As the potential fall within the plasma may be taken as negligible, the edge of the plasma forming a boundary of the cathode sheath is approximately at the same potential as the anode. Hence the plasma may be regarded as an extension of the anode towards the cathode surface. Within the cathode sheath, (9–7) is appropriate with $K \neq 0$. However, if we regard the contribution to the current of the electrons and ions formed by collision as negligible, i.e. $K \doteqdot 0$, then

$$J_- = \frac{\sqrt{2}}{9\pi}\,\frac{e}{m}\,\frac{E^{3/2}}{d^2} \qquad . \quad . \quad . \quad (9\text{–}8)$$

Referring to Fig. 9–3 it will be noted that except for small currents the volt drop across a gas-filled valve is almost independent of the value of the current. Hence if K is small then $J \propto E^{3/2}/d^2$. As E is independent of J_- then the latter must increase by virtue of a decrease in d. In practice, as J increases, the plasma, which is practically at the same potential as the anode, moves nearer to the cathode, thus reducing *d* and making possible a current increase.

VALVE VOLT DROP

From previous statements and Fig. 9–3 it is evident that gas-filled valves only conduct inappreciable currents while the anode voltage is initially below the ionization potential of the gas filling. At a voltage approximately equal to the ionization potential, relatively large currents commence to flow, this potential being generally referred to as the ignition or striking potential. From Fig. 9–3, however, it will be noted that once this potential has been attained the drop across the valve falls to some lower figure, at which it tends to remain as the current is increased. It is found that the volt drop across a gas-filled valve is frequently lower than the ionization potential and may even be lower than the lowest excitation potential. As some electrons may be emitted from the cathode with velocities corresponding to a few electron-volts, it is only necessary for such electrons to fall through a voltage equal to the difference in the ionization potential and that corresponding to the emission velocity for ionization to occur. If excitation to metastable states occurs, then cumulative ionization is possible, and in this case the volt drop across the valve may be very little higher than the lowest excitation potential of the gas.

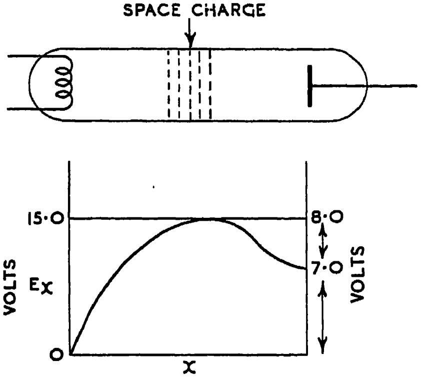


Fig. 9–4

In some cases the volt drop may be even lower than the lowest excitation potential. This is due to a positive space-charge forming between anode and cathode of such magnitude as to reverse the potential difference between itself and the anode. This condition is illustrated by Fig. 9–4, from which it will be noted that between the cathode and space-charge a potential difference exists sufficient to produce ionization. Although the field beyond the space-charge is reversed, electrons pass through this by virtue of the kinetic energy they have acquired in the accelerating field between the cathode and space-charge.

INFLUENCE OF GAS PRESSURE

We have already seen that the quantity of gas has an important influence on the behaviour of a gas-filled diode and we shall here pursue this aspect of the valve still further. Assuming that the gas density is such as to secure a low cathode-anode potential drop, variations in the gas density about this point have an important effect on the four following characteristics: (1) the maximum inverse voltage that may be withstood, (2) the maximum current that may be carried, (3) the cathode-anode drop, and (4) the rate of evaporation of the cathode. With regard to (4), if the gas density is sufficiently increased, it reduces the rate of evaporation of the cathode and thus permits the operation of the cathode at a higher temperature and emission efficiency. This is a practice which is, of course, followed with the so-called ½-watt gas-filled lamp. In the case of the original Tungar rectifier, this employed a thoriated tungsten filament operating in argon at a pressure of approximately 5 cm. of Hg. One disadvantage of a high gas pressure is that it tends to cause the discharge to concentrate on a small area of the cathode, resulting in a burn-out. In fact, if once the discharge is started it will be found to continue if the cathode-heating current is cut off. In Chapter XIX it will be found that this arc-heating of the cathode is employed to advantage in certain luminous discharge lamps.

Apart from the desirability of a high gas density to reduce filament evaporation, it is undesirable in that it reduces the inverse peak voltage that a tube will withstand. The gas pressures employed in practice are such that the inverse breakdown potential lies to the left of the minimum of a curve such as that of Fig. 2–6, and also follows Paschen's Law. Hence if the gas density and pressure are increased, the breakdown potential and inverse peak voltage decrease. It follows that if the gas pressure is too high, the tube may break down during the inverse half-cycle and a glow or arc occur. Under these circumstances the tube fails to rectify and may be destroyed.

Apart from its detrimental effect on cathode evaporation, an attempt to increase the inverse peak voltage by reducing the gas pressure may be accompanied by a further detrimental effect. Should the gas pressure be too low, there may be insufficient gas molecules present to provide a sufficiency of positive ions to neutralize the negative space-charge when the tube is carrying its rated current. This leads to an excessive volt drop and disintegration of the cathode by positive ion bombardment. It is, of course, practicable to avoid this effect by de-rating the current of the tube as the gas pressure is reduced.

As the density of mercury vapour in the presence of its liquid depends upon temperature, it is evident that the foregoing characteristics when referred to a mercury-vapour tube largely depend upon the temperature of the tube. Hence, as the temperature of the tube increases, the maximum inverse voltage and tube drop decrease while the maximum anode current increases. The relation between mercury-vapour pressure and temperature is shown by Fig. 9–5, the effect of temperature on the characteristics of a mercury-filled diode being shown by Fig. 9–6. It is evident from the latter that such tubes must be operated between fairly close limits of temperature, for at high temperatures low inverse voltages and cathode vaporization occur, while at low temperatures high volt drop and low current ratings result. Thus, desirable operating limits are about 20° C. to 60 °C. The temperature of a valve is, of course, related to the ambient temperature and the valve losses. The latter comprise those of the cathode and the valve drop. The cathode heating losses are such that, generally speaking, even at under no-load conditions, the valve temperature is well above the low limit already quoted. For example, the temperature rises of

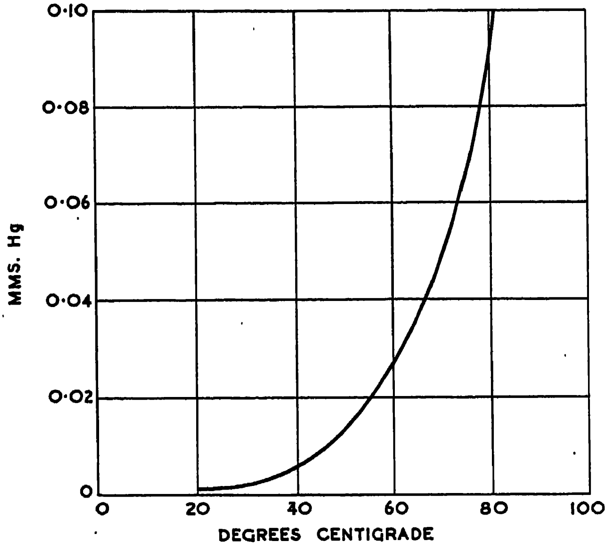


FIG. 9–5

the valves shown by Figs. 13–6 and 13–7 are, respectively, 70° C. and 14° C. due to cathode losses alone.

In tubes in which the gas filling is one of the inert gases, such as neon, argon, etc., there is no variation of gas density with temperature, and variations in tube characteristics, such as those which take place with mercury-filled tubes, do not occur. However, the gas gradually disappears, due to clean-up, with a modification in the performance of the tube due to this cause. In high-voltage tubes the gas pressure must be maintained between narrow limits and hence in such tubes mercury vapour only is employed. In low-voltage tubes, gas pressure is not of such importance and sufficient gas may be initially admitted to compensate for such clean-up as occurs during the useful life of the tube.

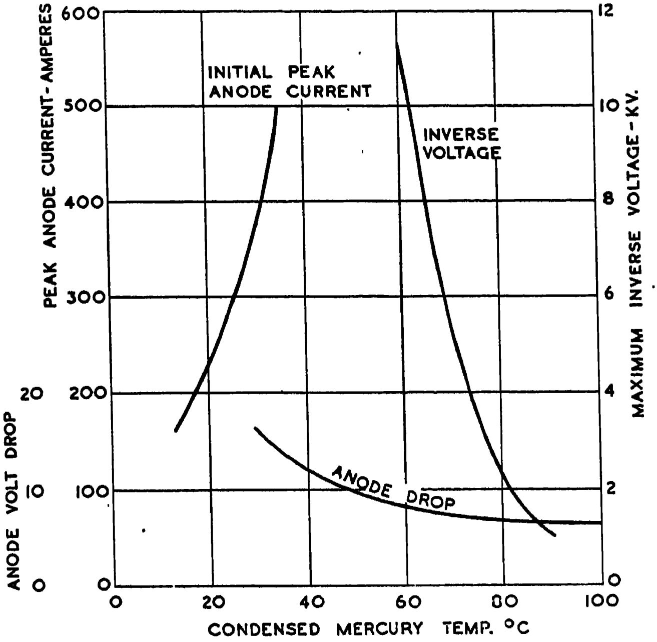


Fig. 9-6

From what has previously been said, it is evident that where high inverse voltages must be withstood it is necessary to work at low gas pressures and, generally, de-rate the current output. However, in addition to this, further steps may be taken which have a beneficial effect.

At the moment the inverse half-cycle commences there are in the vicinity of the anode the positive ions of the plasma. As the anode becomes negative, some of these are swept out of the plasma to the anode and thus may form a glow discharge which, if circumstances are unfavourable, can develop into an arc. Now the glow current is proportional to the area of its cathode, and hence the area of the anode in the gas-filled diode is not made larger than is absolutely necessary. In considering the glow discharge we saw that it is probably the increasing density of the glow current which leads to an arc discharge. This density varies as the square of the gas density and hence a low gas density leads to low glow current density. A further method of increasing the maximum inverse voltage is to enclose the anode in an arm attached to the side of the tube. This encourages a loss of electrons and ions to the wall of the tube, thus increasing the resistance to the initiation of a glow discharge. The placing of anodes in arms is, of course, a very common one with glass-bulb mercury-arc rectifiers and is partly done for the same purpose.

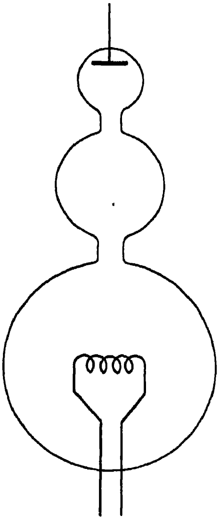


FIG. 9–7

Where very high inverse voltages must be withstood, up to 100 kV, mercury vapour is always employed, maintained at a low temperature. Naturally, the filament vaporizes rapidly, and to obtain a useful life it must be heavily constructed. This means what would normally be an excessive filament consumption for a given current through the tube. The gas lost due to vaporization is, of course, continually replaced by the liquid mercury. A form of tube employed for high inverse voltages is shown by Fig. 9–7. It will be noted that the anode and cathode are placed in separate spaces with a condensing chamber between them. This leads to different gas pressures round the anode and cathode. The cathode space contains liquid mercury which continually vaporizes because of the heating in this space produced by the cathode and cathode sheath losses. After condensation, the mercury flows back to the cathode space under the influence of gravity. With this arrangement, the gas pressure round the cathode is relatively high, resulting in a relatively low degree of cathode vaporization, while, due to the condenser, the pressure in the anode space is relatively low. Thus, a high inverse voltage may be sustained without undue cathode vaporization.

Cathodes in Gas-filled Valves

Because of the low voltage drop of the gas-filled diode, the current-carrying capacity of these tubes is not limited by anode heating. Hence the total emission of the cathode is of greater importance than with a vacuum tube. However, due to the presence of positive ions, a possibility exists of cathode disintegration, or atomization, as it is sometimes termed, caused by positive ion bombardment. As practically the whole tube drop is concentrated immediately in front of the cathode, it follows that practically every ion which strikes the cathode does so with an energy which corresponds to approximately the entire anode-cathode potential difference. It therefore appears possible that this bombardment may result in the destruction of the cathode, particularly if this is of the coated type, as it usually is in gas-filled valves. Experiment has shown, however, that the cathode is but slightly affected provided the ionic energy (and hence anode voltage) and instantaneous current do not exceed certain limits. As the anode voltage is increased, due, say, to a reduction in gas density, the energy of the bombarding ions increases, but is ineffective until what is known as the disintegrating voltage is reached. The value of this is 22 volts for mercury vapour, 25 volts for argon, and 27 volts for neon. Hence the gas density must be such that these voltages are not exceeded if a reasonable cathode life is to be realized.

With regard to current limitations, disintegration will occur if the peak current through the tube exceeds the total thermionic emission of the cathode. Should the peak current exceed the total emission, then the difference must be produced by ionic bombardment of the cathode, i.e. the virtual formation of a cathode spot. This results in an increased temperature of the cathode and its rapid vaporization and destruction. In view of the foregoing, a gas-filled valve is often given two current ratings: (1) an average rating based on temperature rise, and (2) a maximum instantaneous rating based on cathode life. For example, the average rating of the B.T.H. BD12 valve is 33 amp. and the maximum instantaneous current, 100 amp.

In tubes employing directly-heated cathodes it is obvious that one end of the filament is at a higher potential than the other. Thus the anode to cathode potential difference is greater for the negative end of the filament than for the positive. This is of no consequence provided the potential difference is less than the disintegration voltage. Hence if the arc drop in a mercury tube is 10 volts the filament voltage should not exceed 12 volts. If the filament voltage is alternating, then its r.m.s. value should not exceed $12/\sqrt{2} = 8{\cdot}5$ volts. The value adopted in practice is frequently 2 volts and seldom exceeds 5 volts. Incidentally, the latter figure is that for the BD12 valve referred to above. A further possibility with directly-heated valves is the striking of an arc from one end of the filament to the other and this again limits the filament-heating voltage that may be employed. With indirectly-heated cathodes the foregoing restrictions are not applicable and filament voltages of 110 volts are sometimes used.

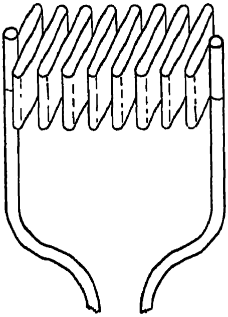


FIG. 9–8

CATHODE EMISSION EFFICIENCY

In the case of gas-filled valves, the extension of the plasma down to within a fraction of a millimetre of the cathode surface permits a profound modification in cathode design, resulting in an enormous increase in cathode efficiency compared with that obtainable with vacuum tubes. In the latter the filament or cathode must be directly exposed to the anode and the distance between them kept to a low figure. With a gas-filled valve, the filament may be placed in a cavity at a considerable distance from the anode, and the plasma will penetrate this cavity, giving a virtual anode at a minute distance from the cathode surface. Thus, the cathode may be so shielded that heating losses are greatly reduced without electrons experiencing any difficulty in reaching the anode in spite of the shielding. For example, by corrugating the cathode, as shown by Fig. 9–8, the electron-emitting surface remains unaffected, but the heat-radiating surface is greatly reduced because of heat reflexion between adjacent parts. Thus, a given temperature may be attained for a smaller current than is necessary with the same cathode uncorrugated, with a consequent gain in efficiency. A typical heat-shielded cathode is shown by Fig. 9–9. In this, heat-shielding cylinders are placed round the cathode, the emitting surface of which is increased by the vane or fin-like structure as shown. The cylinders reflect the heat internally, while the vanes add greatly to the electron-emitting area but little to the heat-radiating surface. Cathodes of this form can give an emission efficiency as great as 1·5 amp. per watt at

1000° K. compared with only 100 mA per watt for a plain oxide-coated cathode employed in a gas-filled tube.

A precaution which must be observed with gas-filled tubes is that the cathode be allowed to attain its normal operating temperature before passing current in the anode circuit. If this is not observed, the tube drop will exceed the disintegration voltage in order to produce by ionic bombardment of the cathode the difference between the total emission of this and the current demanded. Thus the heating time of the BD12 is five minutes.

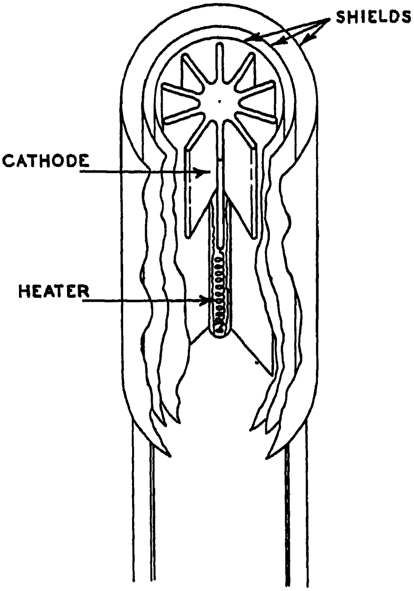


Fig. 9–9

Anode Temperature

Although the output of a gas-filled tube is not limited by anode heating, the anode temperature must not be too high. In due course the anode (and any other electrodes) become coated with a layer of material (deposited from the cathode) of low-work function. In order that any emission from the anode be kept to an inappreciable amount it is necessary that the anode temperature should not exceed about 400° C. for metal and 200° C. for graphite.

The Gas-filled Thermionic Triode: Thyratrons

By adding a grid to a gas-filled diode a valve is produced, termed a thyratron or gas-discharge triode. A thyratron functions in the following manner. Providing the valve has not struck, the field produced by a negative grid potential can so affect that due to a positive anode potential that the striking potential is dependent upon both grid and anode voltages. As the grid potential is slowly changed from negative towards positive, a critical grid voltage is reached, dependent on the anode potential, at which the tube strikes and a plasma is formed. From this point the grid exerts no further appreciable effect, the valve potential falls to between 10 to 20 volts, and the anode current must be limited by an external resistance to avoid destruction of the tube. It will be appreciated that once ignition has occurred, the behaviour of a thyratron is practically identical with that of the thermionic gas-filled diode previously discussed. After a plasma has formed, no matter how negative the grid may be made, the discharge cannot be terminated except by reducing the anode voltage below the arc value or reversing it. Should the arc be extinguished, then after a short period, termed the de-ionization time, the grid, if sufficiently negative, regains control until either the anode voltage is sufficiently increased or the grid negative potential sufficiently reduced for restriking to occur.

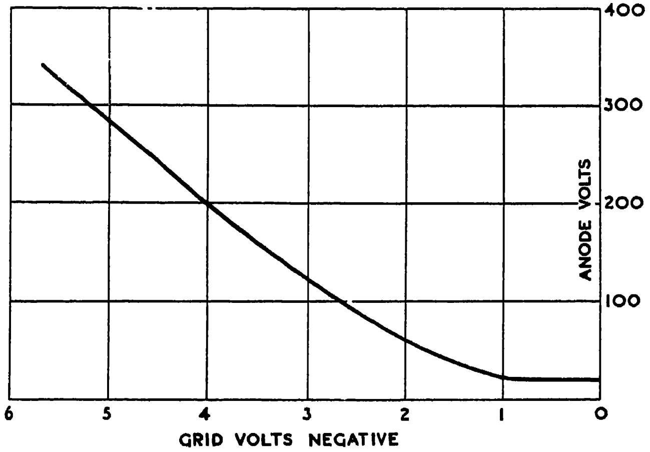


FIG. 9–10

THEORY OF STRIKING: GRID CURRENT

A characteristic curve for a thyratron, Cossor GT4B is shown by Fig. 9–10. This, obviously, shows the relationship between anode and grid potentials for striking, the relationship depending on the geometrical arrangement of the electrodes. The ratio of the anode to grid potential at any point of the characteristic is known as the control ratio and for the case shown is approximately 50 to 1 for points removed from the origin. Before the tube strikes, the electron current which escapes from the cathode is very small, due to the negative grid charge returning most of the emitted electrons to the cathode. Thus, only those electrons emitted from the cathode with exceptionally high velocities will succeed in passing through the grid to the anode. From the few positive ions which are formed some will be collected by the grid forming a positive-ion current to this, the current being conventionally termed a negative grid current. In addition to this, because of the initial velocities of the electrons and the cathode-grid contact potential difference, the grid may also collect electrons giving rise to a positive grid current. In practice the resultant grid current will be the sum of these two components. Although these two components predominate, other possible components which may be mentioned are—

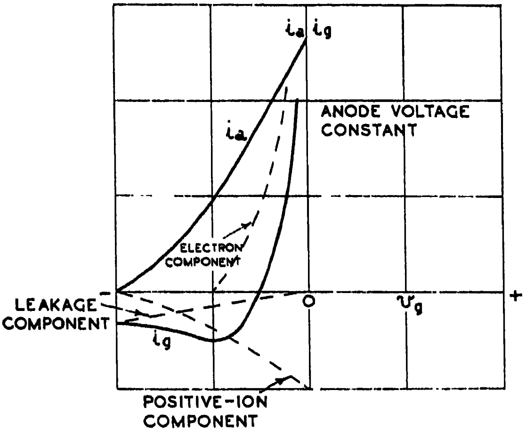


FIG. 9-11

(1) Current due to secondary emission from the grid.
(2) Current due to photo-electric emission from the grid.
(3) Current due to thermionic emission from the grid.
(4) Current due to leakage along insulation-resistance paths.

Because of the Maxwellian distribution of velocities amongst the electrons, it follows that the pre-striking grid current will gradually diminish as the negative potential of the grid is increased.

The various currents in a thyratron prior to striking are shown by Fig. 9-11. The abscissa is the grid potential and the ordinate, to two different scales, the anode and grid currents. Two scales are necessary as the anode current may be of the order of milliamperes and the grid current microamperes. i_a and i_g are, respectively, the anode and grid currents, the positive and negative components of the latter being shown. The positive-ion current is a function of the rate of ionization in the valve and is proportional to the anode current and the gas density. It follows that the positive-ion current (negative grid current) must be zero when the anode current is zero, and that the former must be an image of the latter.

GRID ACTION BEFORE STRIKING

In cases where the anode voltage of a thyratron is low it may be necessary to apply a positive potential to the grid before striking will occur. With a low anode potential, α, i.e. the number of electrons and ions formed per primary electron per centimetre of path, is small because of the low field strength X. Thus, a relatively high current is necessary to form a plasma and this means that to attain such a current the grid must be made positive. In these circumstances the grid current will be positive at the moment of striking, thence changing rapidly to negative. This condition is shown by Curve 1, Fig. 9–12. For higher anode voltages α is higher, less current is necessary to form a plasma, and the tube may strike for negative grid voltages and relatively low grid currents. Such conditions are shown by Curve 2, Fig. 9–12, where negative grid current is flowing at the moment of striking which, however, rapidly increases after the tube has struck. In the majority of instances the pre-striking grid current can be neglected and a thyratron regarded as a purely voltage-operated device. Nevertheless, in certain cases it may be necessary to include a resistance in the grid circuit where it is desired to limit grid current, for, with some grid-controlled apparatus, the grid current may amount to as much as 500 mA.

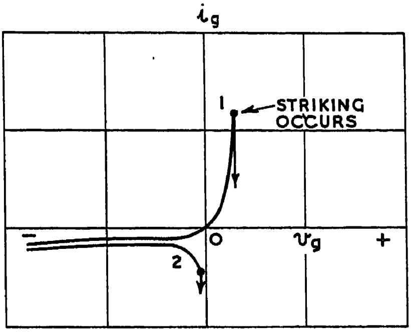


FIG. 9–12

GRID ACTION AFTER STRIKING

It has previously been stated that once striking has occurred the grid exerts no further control in a thyratron, no matter how negative it may be made. We must now examine this statement and the reason for it in greater detail. After the discharge has occurred, a positive-ion sheath forms at the cathode, the remainder of the discharge path consisting of a plasma in which the concentrations of positive ions and electrons are approximately equal.

Because of the low field strength within the plasma the potential of the latter is, approximately, the same as that of the anode. In fact, the plasma may be regarded as an extension of the anode down to almost the cathode surface. When the grid is made negative it repels electrons and attracts positive ions, and thus a positive-ion sheath is formed round the grid. When conditions become stable, the positive-ion space-charge surrounding the grid is equal to the negative grid charge produced by the applied grid-cathode voltage. Hence the net charge within the boundary of the positive-ion sheath is zero and the plasma is unaffected by the presence of the negative grid. In consequence, conduction occurs through the grid spaces unaffected by the grid potential. The various conditions within the thyratron after striking are indicated by Fig. 9–13.

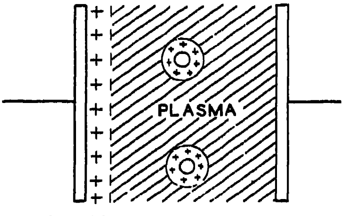


Fig. 9–13

The foregoing discussion is dependent on the assumption that the thickness of the sheaths surrounding the grid wires is small compared with the distance between the wires. Actually the sheaths may be considered as increasing the virtual diameter of the grid wires and under some circumstances this effect is of importance. As the current through the sheath to the grid is carried by positive ions only, this current may be regarded as a positive-ion space-charge-limited current similar to the electron current in a vacuum valve. If the sheath thickness is small compared with the diameter of the grid wire, the space-charge equation for parallel planes may be applied with appropriate modification. Thus, in the present case, (8–5) becomes

$$i_{+} = \frac{2{\cdot}33 \times 10^{-6}}{\sqrt{1839M}} \frac{E_{ga}^{3/2}}{d^2} \quad . \quad . \quad . \quad (9\text{–}9)$$

where i_{+} is the positive ion current per square centimetre of grid surface, M the atomic weight of positive ions, E_{ga} the potential difference between grid and anode, and d is the thickness of the positive-ion sheath. Actually E_{ga} should be the voltage across the sheath, i.e. the potential difference between grid and plasma. As, however, the potential of the latter is practically the same as that of the anode, E_{ga} may be taken as stated.

An important difference between the space-charge-limited electron current in a vacuum valve and the similar positive-ion current in the thyratron grid sheath must now be noted. In the former the current, of course, varies with the cathode-anode potential difference, whereas in the latter, providing the grid is more than a volt or two negative, the current tends to be constant and independent of the magnitude of the grid voltage. This is because the positive-ion current to the grid is due to the *random* motion of the ions in the plasma rather than to a directed motion. In fact, after striking has occurred, the behaviour of a thyratron grid is exactly similar to a negative probe. Hence as the grid voltage is varied i_+ remains constant, from which it follows that the thickness, d, of the positive-ion sheath must vary. Thus, from (9–9)

$$d = \sqrt{\frac{2{\cdot}33 \times 10^{-6} E_{ga}^{3/2}}{\sqrt{1839M}\, i_+}} \quad . \quad . \quad . \quad (9\text{–}10)$$

As the positive-ion density in the plasma is proportional to the anode current, i_+ will vary directly as the latter quantity. Thus the greater the anode current the greater will be the grid current and the smaller the thickness of the grid sheath.

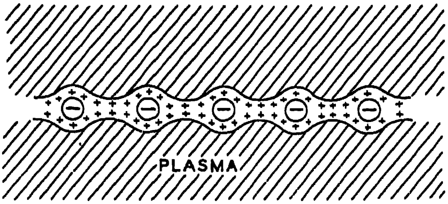


FIG. 9–14

In order to gain some idea of sheath thickness we may select typical values for the various terms in (9–10) and calculate d. Assume a mercury-vapour tube to be under consideration, with a cathode-anode drop of 10 volts, and a negative cathode-grid drop of 5 volts. The random positive-ion current density is taken to be 10^{-3} amp. per cm.^{2} and M, 200·6. Thus $E_{ga} = 15$ and $d = 0{\cdot}015$ cm. or 0·15 mm. Now, although it has been previously stated that once a thyratron has struck the grid is without further control, there is, nevertheless, a special case where the grid may exert an influence on the anode current and even extinguish it. Referring to (9–10) we see that by making E_{ga} large and i_+ small, relatively thick sheaths may be produced, and thus the effective diameter of the grid wires, or grid dimensions, considerably increased. If this process is carried sufficiently far, it is possible to cause the grid sheaths to overlap in the manner shown by Fig. 9–14. As this condition is approached, the anode current will be reduced and, at overlap, extinguished. Now i_+ is proportional to anode current and, therefore, grid control, after striking, can only be effected for low plasma densities as well as high grid potentials.

Considering now the effect of reducing the negative potential of a thyratron operating under normal conditions, as a positive potential is approached some of the electrons within the plasma, because of their random motion, find their way to the grid, thus constituting a positive-grid current. The negative-grid current, while the grid is still negative, remains unchanged, and the resultant grid current is now that due to the ions and electrons. Hence as the grid approaches a positive potential the grid current at first decreases. As the random current density of electrons in the plasma is in excess of that of positive ions, the grid current generally becomes zero while the grid is still negative. As the grid potential changes from negative to positive the electron current increases while the positive-ion current decreases. When the positive grid potential becomes that of the plasma, the positive-ion sheath disappears, and a pure, saturation, electron current results with the plasma extending to the grid surface. If the positive grid potential is continually increased the grid begins to usurp the function of the anode.

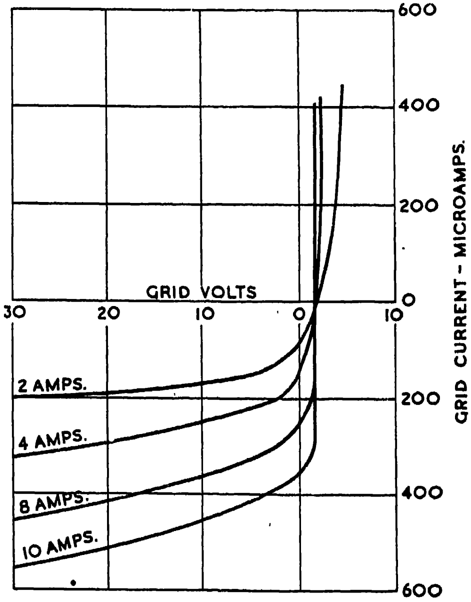


Fig. 9–15

The manner in which the grid current varies with changes in grid potential at different anode currents is shown by Fig. 9–15. It will be noted that an approximately linear relationship exists between grid and anode currents when the grid voltage is held constant. In the vicinity of zero grid voltage the rate of change of grid current is very high and, in practice, it is frequently necessary to include a current-limiting impedance in the grid circuit. This is particularly so when only an inappreciable load must be imposed on the grid voltage source, or when the grid potential is alternating in character and thus may become largely positive at every half-cycle.

THYRATRON GRIDS

The grids employed in thyratrons are markedly different from those in vacuum tubes. Striking occurs when the current density reaches a certain critical value and this may occur at various places in the cathode-anode space. Further, it is possible for the critical current density to be reached at a position outside rather than through the grid spaces. In view of this it is necessary to restrict the anode current to paths which must pass through the grid, and this means that, in gas tubes, the grid must more completely surround either the anode or cathode than is necessary with vacuum tubes. Where the grid surrounds either the anode or cathode alone, the space between the grid and the unshielded electrode will be subjected to the influence of fields produced by charges on the glass walls of the tube. Such charges and their accompanying fields are variable from time to time and result in erratic starting characteristics. Hence it is frequently necessary to extend the grid function to include screening the anode and cathode from the walls of the tube. This, of course, refers to a non-conducting wall, such as glass. When the wall is metal, as with grid-controlled mercury-arc rectifiers, the potential of the wall remains constant and screening of the electrodes from the walls is unnecessary. A typical grid enclosing both electrodes is shown by the thyratron of Fig. 9–16.

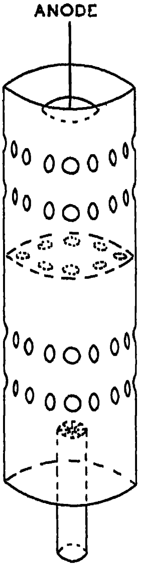


FIG. 9–16

Referring to Fig. 9–16, it will be noted the grid spaces consist of circular perforations. The visible part of the grid forms the shield necessary to neutralize the influence of the tube walls, while the grid proper consists of a perforated horizontal disc mounted within the shield between cathode and anode. As striking occurs due to the pre-striking anode current attaining the critical density at some point, it follows that if the grid possesses a number of perforations, only one of these will be utilized at striking. Thus, such grids frequently possess one opening only, as shown by Fig. 9–17. In this case it will be noted that a shield is employed without perforations, the plasma being thus confined within the shield.

A disadvantage of the shielded grid is that its relatively large dimensions lead to large grid currents and a high inter-electrode capacitance. In order to minimize this, thyratrons with two grids are sometimes employed. These grids are, respectively, termed the shield and control grids, the function of the first being largely to absorb the grid-starting current. This is effected by connecting the grid to a source of fixed potential, when it will tend to intercept most of the positive-ion current. The control grid, which may be a short cylinder, is mounted between the shield grid baffles, as shown by Fig. 9-18, and because of its small area has a low capacitance and only requires a small starting current.

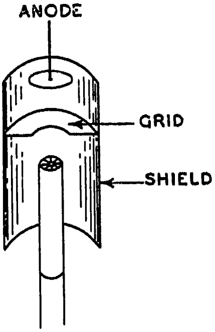


FIG. 9 17

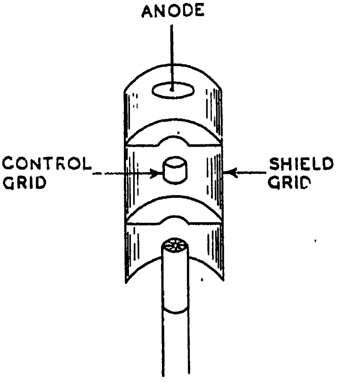


FIG. 9-18

IONIZATION AND DE-IONIZATION TIMES*

In many thyratron applications the speed with which striking may be effected and also the rapidity with which regain of grid control is possible are of great importance. The first case is concerned with the time necessary for formation of the plasma, and the establishment of the anode current to its final value, after the anode or grid potentials are suddenly adjusted to values at which striking occurs. In general this time is exceedingly short, seldom exceeding a few microseconds, and depends on the tube geometry, the electrode voltages, the gas pressure and the circuit in which the thyratron is connected. Evidently the circuit, particularly if inductive, may have a greater effect than the previously-mentioned factors. The de-ionization time depends on the same factors as the ionization time, but is generally longer, and may be of the order of several hundred microseconds. De-ionization depends on the removal of ions and electrons from the discharge path. This may occur by two methods: (1) re-combination within the plasma, (2) diffusion to the electrodes and walls of the tube. The first of these may be neglected, as it is known that the possibility of re-combination rapidly decreases with an increase in the relative speed of the ions, being practically non-existent at energies greater than 0·1 volt. Thus, de-ionization depends on the diffusion of the ions to the electrodes and walls of the tube, where those of opposite sign will neutralize each other. In the case of insulated surfaces, these acquire such a potential that equal numbers of electrons and ions arrive at unit area in unit time, this, of course, being the most effective condition for promoting re-combination and de-ionization. Because of the higher velocity of electrons within the plasma, an insulated surface acquires a potential a few volts negative to the plasma.

---

* "Note on the Ionization and De-ionization Times of Gas-filled Thyratrons," J. C. R. Cance, *J. Sci. Inst.*, Mar. 1946, p. 50.

Referring to (9–10), we may note that as the positive-ion concentration decreases the grid sheath thickness increases. When the concentration is such that overlapping occurs the grid regains control and re-application of a positive anode potential may be made without conduction recurring.

INFLUENCE OF GAS PRESSURE ON THYRATRON CHARACTERISTICS

The operating characteristics of a thyratron after striking has occurred are effected by variations in gas pressure in an exactly similar manner as the characteristics of a diode. In addition, the starting characteristics are also affected by gas pressure in the manner indicated by Fig. 9–19, which concerns a mercury-vapour thyratron. As the temperature is raised the gas density and collision frequency increase. This results in an increase in the efficiency of ionization, which means more ions per primary electron. Thus, the critical current required for plasma formation may be attained at a lower anode voltage, and hence for a given negative grid potential the anode voltage necessary for striking the tube becomes lower as the temperature and gas density are raised. In the case of tubes filled with an inert gas, such as that of Fig. 9–10, variation of the striking characteristics with temperature change does not occur. However, due to clean-up, the gas pressure progressively decreases during the life of the tube and, for a given anode voltage, the negative grid voltage necessary for striking must be reduced as the age of the tube increases.

A further limiting effect of gas pressure in thyratrons is the permissible grid-anode potential difference. It is evident that, if this exceeds a certain value, depending on the gas density, a glow discharge may develop between grid and anode. Should this occur, a positive-ion sheath may form round the grid, causing this to lose control and permitting the main discharge to start. Thus, the grid-anode potential difference is definitely limited and a maximum rating is generally specified for tubes dependent on their temperature. As the grid-cathode potential difference is usually relatively small, it is customary to specify the maximum forward cathode-anode voltage which a thyratron will withstand without a breakdown occurring between anode and grid. This voltage is governed in a similar manner to that of the curve of Fig. 9–6, an increase in temperature and gas density resulting in a reduction in the maximum permissible voltage. Accordingly, the upper limit of the operating temperature of a mercury-vapour thyratron is governed either by the maximum forward anode voltage or the maximum inverse voltage which the tube must withstand. In the case of tubes filled with an inert gas the maximum forward voltage is unaffected by temperature and for the tube of Fig. 9-10 is fixed at 350 volts. Such tubes generally withhold a lower voltage than mercury-vapour tubes, as an excess of gas must be initially admitted to allow for clean-up.

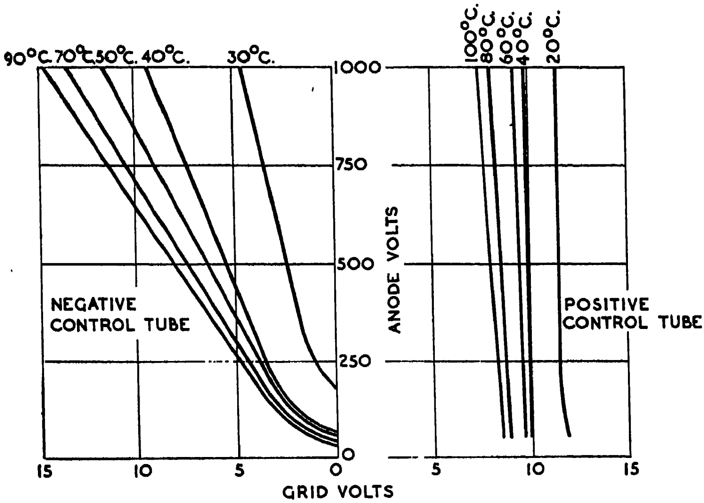


FIG. 9–19

CHAPTER X

HIGH-VACUUM THERMIONIC TRIODES AND MULTI-ELECTRODE VALVES

THE introduction of one or more additional electrodes into a thermionic vacuum diode may profoundly modify its characteristics and certainly vastly increase its field of application. In the case of a diode, for a given filament temperature, the properties of the tube may be expressed by a single characteristic relation $i_a = f(v_a)$, where i_a and v_a are, respectively, the anode current and voltage. In the case of a multi-electrode tube, the current to any particular electrode generally depends on the potentials of all the electrodes relatively to the cathode. In the case of a triode, a third electrode of grid-like structure is interposed between the cathode and anode. If the cathode is a straight filament surrounded by a concentric cylindrical anode, the grid usually takes the form of a widely-spaced wire spiral, concentric with the cathode, occupying a position in the cathode-anode inter-electrode space. With the grid at zero potential the cathode electrons pass through the meshes of the grid and are collected by the anode in a similar manner as with a diode.

Mutual Characteristics

As the anode and grid currents in a triode are, respectively, functions of both the grid and anode potentials, they may be expressed as

$$i_a = f_1(v_a, v_g) \quad . \quad . \quad . \quad . \quad (10\text{–}1)$$

$$i_g = f_2(v_a, v_g) \quad . \quad . \quad . \quad . \quad (10\text{–}2)$$

where v_g and i_g are, respectively, the grid potential and current. The principal function of the grid, however, is not so much to collect current as to control the anode current, and, in practice, means are usually adopted to prevent grid current flowing. Hence we are usually more concerned with f_1 than f_2. Because of the closer proximity of the grid to the cathode than the plate to the cathode, the controlling effect of the grid on the anode current is generally far greater than that of the anode. A result of this is that grid-cathode potentials are usually much smaller than are anode-cathode potentials. As the relation of the anode current to the grid and anode potentials can only be expressed by simple functions over a limited range, it is customary to express f_1 (and f_2) by a family of curves. A particularly important set of curves is that showing the relation between the anode current and grid potential for various constant values of the anode potential. Such curves are shown by Fig. 10–1 and are known as *mutual characteristics*. It is found that (10–1) may be replaced by

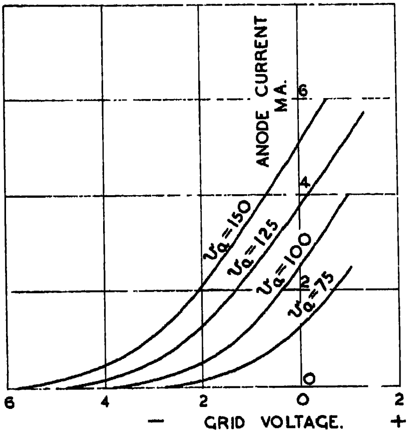


Fig. 10–1

$$i_a = k(v_a + \mu v_g)^n \quad . \quad . \quad . \quad (10\text{–}3)$$

where k, μ, and n are constants. It is evident that μ expresses the relative influence of the grid potential on the anode current as compared with the anode potential. Hence μ is known as the *amplification factor*. n is approximately equal to 3/2 and this indicates that (10–3) is an expression of the three-halves power law. We shall now attempt to justify (10–3) and determine the factors on which μ depends.

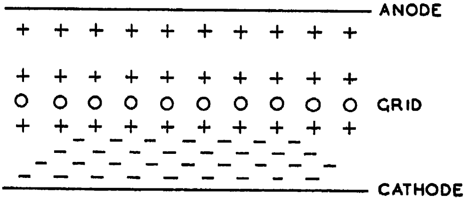


Fig. 10–2

Referring to the triode structure of Fig. 10–2, assume that both grid and anode potentials are positive with respect to the cathode. Then the combined positive charges on grid and anode will be equal to the negative space-charge. Thus, the grid and anode combined are equivalent to the anode of a diode and it is possible to replace any triode by an equivalent diode, the anode potential of which simulates the combined effects of the grid and anode of the triode. The anode of the equivalent diode may be considered as situated at the same position as the grid of the triode it represents, these conditions being indicated by Fig. 10–3.

It is now desired to find a potential v at which the anode of the diode must be maintained in order that it may simulate the triode when the grid and anode potentials of the latter are, respectively, equal to v_g and v_a. This means that v must be such as to make the charge on the diode anode equal to the combined charges on the grid and anode of the triode. Each of the systems of Fig. 10–3 may be regarded as a capacitance and, in the case of the triode, the grid and anode may be considered as a single electrode. The capacity of the diode will be denoted by C_{14} and those of the triode by C_{12} and C_{13}. The charges on the two systems will be equal when

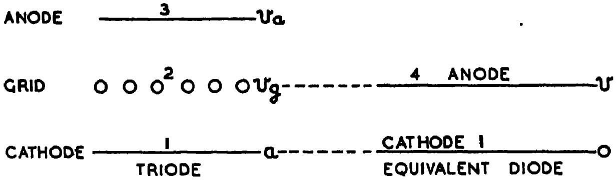


FIG. 10–3

$$C_{12}v_g + C_{13}v_a = C_{14}v$$

from which

$$v = \frac{C_{12}}{C_{14}}\left(v_g + \frac{C_{13}}{C_{12}}v_a\right)$$

this giving the anode voltage of the equivalent diode which will simulate the combined effects of the grid and anode of the triode. In practice, C_{12} frequently differs but slightly from C_{14} and in this case we may write

$$v = v_g + \frac{C_{13}}{C_{12}}v_a$$

Thus, the diode anode potential necessary to simulate the triode is equal to the grid potential of the latter plus a fraction of its anode potential. The quantity C_{13}/C_{12} is known as the penetration factor (denoted by D) and may be regarded as a measure of the extent to which the field from the anode penetrates through the grid to the cathode. The reciprocal of D is termed the amplification factor and is denoted by μ. It is evident that if C_{12} and C_{13} can be determined from the geometry of the electrode system, then D and μ may be evaluated.

If the value of v is substituted in the diode formula of (8–5) there results

$$i = i_a + i_g = 2{\cdot}33 \times 10^{-6} \frac{(v_g + Dv_a)^{3/2}}{d^2} \text{ amp./cm.} \qquad (10\text{–}4)$$

or

$$2{\cdot}33 \times 10^{-6} \frac{(v_g + v_a/\mu)^{3/2}}{d^2} \text{ amp./cm.} \quad . \quad . \qquad (10\text{–}5)$$

where i is the total cathode current. From these formulae the importance of D and μ is at once apparent, for it is seen that control by the grid of the current leaving the cathode is μ times as effective as control by the anode. The expressions given by (10–4) and (10–5) justify (10–3), for it is evident that (10–5) may be written

$$i = i_a + i_g = \frac{2{\cdot}33 \times 10^{-6}}{\mu^{3/2}d^2} (v_a + \mu v_g)^{3/2}$$

We have so far considered the potentials of both the grid and anode to be positive with respect to the cathode. However, providing $(v_g + Dv_a)$ is positive, current will flow in the equivalent diode even if one of the electrode potentials is negative. If one electrode is negative, then the entire cathode current passes to the positive electrode. In this case (10–4) may be written

$$i = i_a = \frac{2{\cdot}33 \times 10^{-6}}{\mu^{3/2}d^2} (v_a + \mu v_g)^{3/2}$$

on the assumption that v_g is negative. In practice the usual function of the grid is to control the current to other electrodes rather than collect it, and hence it follows that the absolute value of v_g is generally small compared with v_a. When both grid and anode are positive, both collect current. However, the electrons acquire high velocities in the inter-electrode space and the majority of these pass through the meshes of the grid and are collected by the anode. Those which are collected by the grid consist, principally, of electrons which leave the cathode from positions immediately opposite grid wires. Thus, approximately, the ratio of i_g to i_a is the same as the ratio of the projections of the grid and anode as viewed from the cathode.

From the foregoing study of the triode it is evident that grid control of the anode current is greater than that of the anode because of the proximity of the grid to the cathode. The grid is greatly inferior to the anode as a collector of current because of the high velocities of the electrons and its small projected area. Thus the main function of the grid is to control the current, while that of the anode is to collect it.

TRIODE CHARACTERISTICS

In practice the three-halves power law has little application to triodes and multi-electrode valves, partly because of the small range over which it is applicable, and partly because of difficulties of employing it for quantitative purposes. Referring to Fig. 10–1, as saturation is approached it will be noted that these characteristics tend to be linear over an appreciable part of their length and, in practice, conditions are usually arranged so that the linear portion is the effective working part of the characteristic. Hence it is evident that here the three-halves power law is inapplicable and that the characteristics of the tube must be represented by a series of curves as shown by Fig. 10–1. With the assistance of Fig. 10–4, three important triode constants will now be defined.

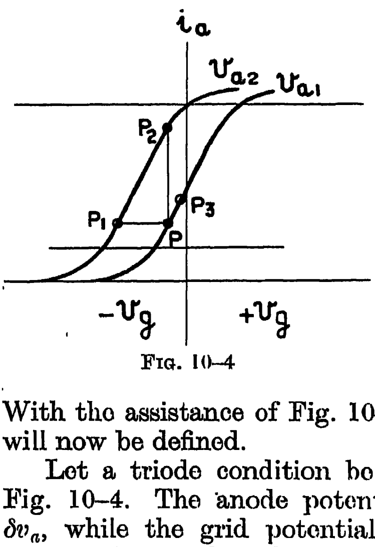


FIG. 10–4

Let a triode condition be such that its operating point is P, Fig. 10–4. The anode potential is now increased by an amount δv_a, while the grid potential is simultaneously decreased by an amount δv_g, so that the operating point is moved to P_1, the anode current remaining unchanged. Then, when δv_a and δv_g become indefinitely small, the fraction $\delta v_a/\delta v_g$ is the *amplification factor* of the triode denoted by μ. Now with the initial point again at P let the anode voltage be increased, while v_g remains constant, so that the operating point moves to P_2 and the anode current increases by a small amount δi_a. Again, when δv_a and δi_a are indefinitely small, the fraction $\delta v_a/\delta i_a$ is termed the *anode impedance* or *slope resistance* of the triode. It must be emphasized that this resistance is different from that obtained by the application of Ohm's Law. In the latter case the resistance is, of course, v_a/i_a, whereas the slope resistance is the resistance offered by the triode to a *change* in current. This latter resistance is denoted by the symbol ρ. There remains one constant to be defined. Commencing from the point P with the anode voltage maintained constant, let the grid voltage be increased so that the anode current increases from P to P_3. If the respective increases are δv_g and δi_a, then, when these are indefinitely small, the fraction $\delta i_a/\delta v_g$ is termed the mutual conductance of the triode and is denoted by g_m. Collecting results we have

$$\mu = \frac{\delta v_a}{\delta v_g}$$

$$\rho = \frac{\delta v_a}{\delta i_a}$$

$$g_m = \frac{\delta i_a}{\delta v_g}$$

from which it is evident that $g_m = \mu/\rho$.

The foregoing results may now be applied to the consideration of the variation of the anode current with variations in the grid and anode potentials. Thus

$$\begin{aligned} \delta i_a &= \frac{\delta i_a}{\delta v_a}\,\delta v_a + \frac{\delta i_a}{\delta v_g}\,\delta v_g \\ &= \frac{1}{\rho}\,\delta v_a + g_m \delta v_g \\ &= \frac{1}{\rho}\,(\delta v_a + \mu \delta v_g) \end{aligned}$$

Now, if conditions are such that only the straight portions of the characteristics of Fig. 10–4 are employed, then μ and ρ are constant and we may write

$$i_a = \frac{1}{\rho}\,(v_a + \mu v_g) \quad . \quad . \quad . \quad (10\text{–}6)$$

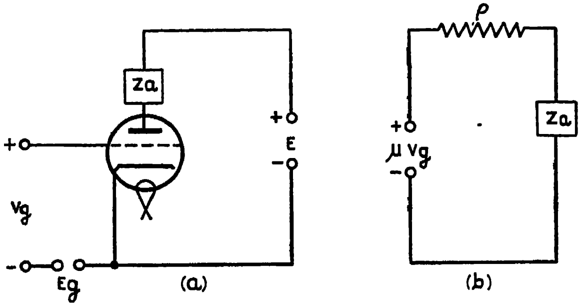
where i_a, v_a, and v_g are understood to refer to *variations* of these quantities from their initial steady values. Comparing this result with that of (10–4), it will be seen that although the total current follows the three-halves power law, the law relating changes of i_a with changes in v_a and v_g is linear.

The initial steady values of i_a, v_a, and v_g are produced by steady d.c. potentials, the latter originating either from batteries or smoothed rectified a.c. sources. The effect of grid current has been ignored in the study of the triode characteristics because, generally speaking, conditions are so arranged that its magnitude is negligible.

The Triode as a Circuit Element

In considering the applications of a triode as a circuit element it may be regarded as a three-terminal network, the terminals being the anode, the cathode, and the grid. Conventionally the cathode terminal is considered as being at zero potential and the potentials of the other two terminals are measured relatively to this. Some circuit properties of the triode will now be considered.

The fundamental circuit of a triode employed as an amplifier or rectifier is shown by Fig. 10–5 (*a*). The steady potentials and anode current are produced by E_g and E, while in the anode circuit is an impedance. Now, from (10–6)



(a) (b)

FIG. 10–5

$$\rho i_a - v_a = \mu v_g$$

or, as E is constant, $v_a + v_z = 0$ and

$$\rho i_a + v_z = \mu v_g$$

Thus, the sum of the instantaneous voltage changes on the anode and Z_a is equal to the change in grid voltage multiplied by the amplification factor μ. From this result it follows that the circuit shown at (*a*) may be replaced by the equivalent one at (*b*); i.e. the actual circuit may be considered as one consisting of ρ and Z_a in series with an e.m.f. μV_g impressed upon it. Hence if μV_g is of sine form we have, employing symbolical notation,

$$\mu V_g = I_a(\rho + R + jX)$$

where $j = \sqrt{-1}$ and $Z = \sqrt{R^2 + X^2}$

It will be noted that although we are dealing with alternating quantities, polarities have been marked in the diagram of (*b*). These polarities are instantaneous ones and indicate the phase-reversing action of the triode. Thus, when the grid potential is increasing, so also is the anode current. It follows that in these circumstances the anode potential must be decreasing and hence the grid and anode alternating potentials are 180° out of phase.

Although the results just derived may be considered correct at audio-frequencies, they are not necessarily so at radio-frequencies. This is because account has not been taken of the various inter-electrode capacities of the triode. These are three in number, namely, C_1, the grid-cathode capacity, C_2, the anode-grid capacity, and C_3, the anode-cathode capacity. Hence the actual circuit should be shown as that of Fig. 10–6 (*a*), where an input impedance Z_g has been included to represent a condition frequently met in practice. In drawing the equivalent circuit the capacitances C_1 and C_3 will be assumed as included in Z_g and Z_a, respectively, while it is evident that C_2 will connect G and A in Fig. 10–6 (*b*). Hence the equivalent circuit when the inter-electrode capacities are considered is given by Fig. 10–6 (*b*).

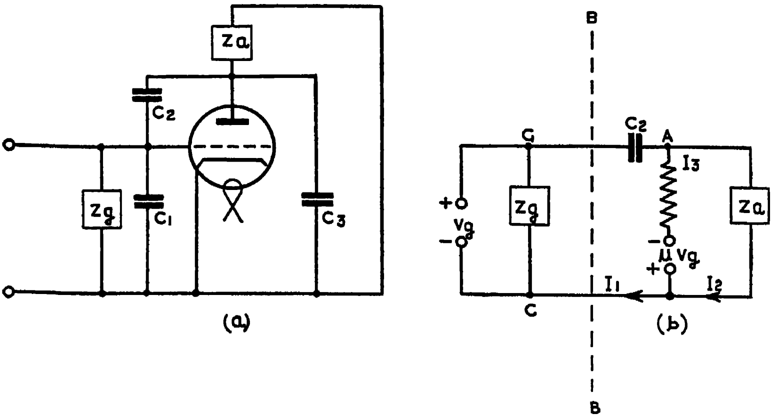


Fig. 10–6

Triode Input Impedance

Consideration of Fig. 10–6 shows that, unlike the circuit of Fig. 10–5, where the input impedance is infinite, the input impedance of a triode is finite because, not only is an impedance $\mathbf{Z}_g$* directly connected to its terminals, but the grid circuit is coupled to the anode circuit by C_2. Denoting the impedance of the circuit to the right of BB by $\mathbf{Z}_i$ we have

---

* Bold face type is here employed to denote complex quantities.

$$\mathbf{Z}_i = V_g/I_1 \quad . \quad . \quad . \quad (10\text{–}7)$$

and if I_1 is determined $\mathbf{Z}_i$ may be found.

From the network of Fig. 10–6 (b) we have, employing Kirchhoff's Laws,

$$I_1 - I_2 - I_3 = 0 \quad . \quad . \quad . \quad . \quad (10\text{–}8)$$

$$-I_1 j/pC_2 + I_3\rho = V_g + \mu V_g \quad . \quad . \quad (10\text{–}9)$$

$$-I_1 j/pC_2 + I_2\mathbf{Z}_a = V_g \quad . \quad . \quad . \quad (10\text{–}10)$$

where $p = 2\pi f$, f being the frequency of V_g.

From (10–8)

$$I_3 = I_1 - I_2$$

and substituting in (10–9)

$$-I_1 j/pC_2 + (I_1 - I_2)\rho = V_g + \mu V_g$$

or

$$I_1(\rho - j/pC_2) - I_2\rho = (1 + \mu)V_g \quad . \quad . \quad (10\text{–}11)$$

From (10–10)

$$I_2 = \frac{V_g + I_1 j/pC_2}{\mathbf{Z}_a} \quad . \quad . \quad . \quad (10\text{–}12)$$

and putting in (10–11) we have

$$I_1(\rho - j/pC_2) - \frac{\rho}{\mathbf{Z}_a}(V_g + I_1 j/pC_2) = (1 + \mu)V_g$$

from which

$$I_1 = \frac{(1 + \mu + \rho/\mathbf{Z}_a)}{[\rho - j/pC_2 \,.\, (1 + \rho/\mathbf{Z}_a)]} V_g \quad . \quad . \quad (10\text{–}13)$$

Thus, from (10–7),

$$\mathbf{Z}_i = \frac{[\rho - j/pC_2 \,.\, (1 + \rho/\mathbf{Z}_a)]}{(1 + \mu + \rho/\mathbf{Z}_a)} \quad . \quad . \quad (10\text{–}14)$$

We shall first consider the value of $\mathbf{Z}_i$ at audio-frequencies. Replacing $\mathbf{Z}_a$ by $R_a + jX_a$, it is evident, if $\mathbf{Z}_a$ is zero,

$$\mathbf{Z}_i = -j/pC_2$$

and the input impedance corresponds to a capacitance

$$C_1 + C_2$$

If $\mathbf{Z}_a$ tends to be extremely large then

$$\mathbf{Z}_i = \frac{\rho - j/pC_2}{1 + \mu}$$

or, as $1/pC_2$ is invariably large compared with ρ,

$$\mathbf{Z}_i = -\frac{j}{(1+\mu)pC_2} = \frac{1}{jp(1+\mu)C_2} \text{ approx.}$$

Hence, in this case, the input impedance corresponds to a capacitance

$$C_1 + (1+\mu)C_2$$

Of course, in the latter case the input impedance actually consists of a resistance in series with a capacitance.

TRIODE INPUT IMPEDANCE AT RADIO-FREQUENCIES

Considering now the value of $\mathbf{Z}_i$ at radio-frequencies, three cases of practical interest may be distinguished: (1) when the anode impedance is a pure resistance ($X_a = 0$), a pure inductance $R_a = 0$, X_a positive), or a pure capacitance ($R_a = 0$, X_a negative). For case (1)

$$\begin{aligned}\mathbf{Z}_i &= \frac{[\rho - j/pC_2 \,.\, (1 + \rho/R_a)]}{(1 + \mu + \rho/R_a)} \\ &= \frac{R_a\rho}{R_a(1+\mu)+\rho} - \frac{j}{pC_2}\left[\frac{R_a + \rho}{R_a(1+\mu)+\rho}\right] . \qquad (10\text{–}15)\end{aligned}$$

From which it is evident that $\mathbf{Z}_i$ consists of a resistance in series with a capacitance. If R_a is very high then the effective resistance is $\rho/(1+\mu) = 1/g_m$, approximately. Also the input capacitance is approximately equal to $C_2(1+\mu)$. At very high frequencies, the reactance of C_2 tends to zero and in this case the input impedance of the triode consists of a resistance, $R_a\rho/[R_a(1+\mu)+\rho]$, in parallel with a capacitance C_1.

When the anode impedance consists of a pure inductance,

$$\mathbf{Z}_i = \frac{[\rho - j/pC_2(1 + \rho/jX_a)]}{(1 + \mu + \rho/jX_a)}$$

where X_a is positive.

Writing $a = (1+\mu)$, and rationalizing,

$$\begin{aligned}\mathbf{Z}_i &= \frac{[\rho - j/pC_2(1 + \rho/jX_a)](a - \rho/jX_a)}{a^2 + (\rho/X_a)^2} \\ &= \frac{a\rho - ja/pC_2 \,.\, (1 + \rho/jX_a) - \rho^2/jX_a + \rho/pC_2X_a \,.\, (1 + \rho/jX_a)}{a^2 + (\rho/X_a)^2} \\ &= \frac{a\rho - ja/pC_2 - a\rho/pC_2X_a - \rho^2/jX_a + \rho/pC_2X_a + \rho^2/jpC_2X_a^2}{a^2 + (\rho/X_a)^2}\end{aligned}$$

Collecting real and unreal terms

$$\frac{(a\rho - a\rho/pC_2X_a + \rho/pC_2X_a) - j(a/pC_2 - \rho^2/X_a + \rho^2/pC_2X_a^2)}{a^2 + (\rho/X_a^2)}$$

$$= \frac{\rho\left(1 + \mu - \dfrac{\mu}{pC_2X_a}\right) - j\left[\dfrac{1+\mu}{pC_2} + \dfrac{\rho^2}{X_a}\left(\dfrac{1}{pC_2X_a} - 1\right)\right]}{(1+\mu)^2 + (\rho/X_a)^2} \qquad (10\text{–}16)$$

From this result it will be noted that it is possible for the input resistance to be negative. This means that, under certain conditions, it is possible for the circuit to generate and thus behave as an oscillator. In most cases this is undesirable, and steps must be taken to neutralize the grid-anode capacity to which the possibility of spontaneous oscillation is due. The usual method is to place an internal screen between grid and anode of the valve, in which case the valve is termed a tetrode. This valve is dealt with towards the end of the present chapter. The effect of the screen is to reduce C_2 to a very low value in which event, although R_i may be still negative, the absolute value of the input impedance may be so high as to prevent oscillation occurring.

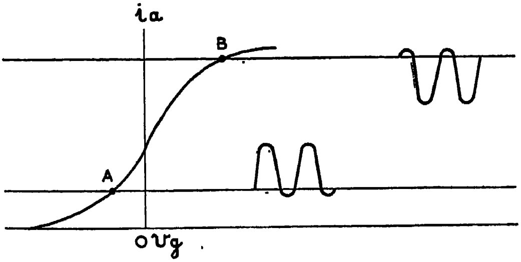


FIG. 10–7

Referring to (10–16), if the anode impedance is a pure capacitance, X_a is negative. In these circumstances it will be noted that the input resistance is always positive and oscillation cannot occur.

The Rectifying Properties of Triodes

Because the grid voltage/anode current and grid voltage/grid current characteristics of a triode are not strictly linear, it is possible to employ triodes as rectifiers. Considering Fig. 10–7, let the anode voltage be such that the operating point A occurs on the curved portion of the characteristic. If an alternating potential, whose mean value is zero, is now applied to the grid the increase in anode current during the positive half-cycles will be greater than the decrease during the negative half-cycles, with the result that an increase in the *mean* anode current will occur. If the initial operating point were at B, then, of course, a decrease in anode current would occur.

To deal with this matter quantitatively, let the relation between grid voltage and anode current be given by

$$i_a = i_Q + av_g + bv_g^2 \quad . \quad . \quad . \quad (10\text{–}17)$$

If now a sinusoidal e.m.f., V_m sin pt, is applied between grid and cathode, we have

$$i_a = i_Q + aV_m \sin pt + bV_m^2 \sin pt$$

and the mean value of this taken over a period is

$$I_a = i_Q + bV_m^2/2$$

Hence it is evident that an increase in mean current of value $bV_m^2/2$ occurs.

Rectification of Modulated Voltages

An important application of the rectifying properties of the triode is the rectification of a modulated voltage wave. Such waves usually consist of a radio-frequency oscillation of constant amplitude which is modulated at audio frequency. Included in the anode circuit of the triode are telephones or a loudspeaker the purpose of which is to record the audio-frequency component. Referring to Fig. 10–8, which shows a modulated wave, this may be represented by

$$v_g = V_m(1 + m \sin wt) \sin pt \quad . \quad . \quad (10\text{–}18)$$

where m is the fractional depth of modulation, w the angular frequency of modulation, and p the angular frequency of the high-frequency wave. Applying (10–18) to the grid of a triode, the anode current is given by

$$\begin{aligned} i_a &= i_Q + aV_m(1 + m \sin wt) \sin pt + bV_m^2(1 + m \sin wt)^2 \sin^2 pt \\ &= i_Q + aV_m(1 + m \sin wt) \sin pt \\ &\qquad + \tfrac{1}{2}bV_m^2(1 - \cos 2pt)\left[1 + 2m \sin wt + \frac{m^2}{2}(1 - \cos 2wt)\right] \end{aligned}$$

Now the telephones will not respond to the high-frequency components in this expression and hence these components will be ignored. Rewriting and neglecting terms containing pt,

$$i_a = i_Q + \tfrac{1}{2}bV_m^2\left(1 + \frac{m^2}{2}\right) + bV_m^2 m \sin wt - \frac{bV_m^2 m^2}{4} \cos 2wt$$

From this it will be seen that due to impressing a modulated voltage between grid and cathode the following results are obtained—

(1) An increase in mean anode current equal to $\frac{1}{2}bV_m^2\left(1+\frac{m^2}{2}\right)$.

(2) A periodic variation in anode current of amplitude bV_m^2m at the modulating frequency $w/2\pi$.

(3) A periodic variation in anode current of amplitude $bV_m^2m^2/4$ at frequency w/π.

If the grid voltage/anode current characteristic were linear, no effect would have occurred in the anode circuit. Thus, because of the rectifying properties of the triode, it is possible to produce an audio-frequency current in the anode circuit, although a second harmonic of this is produced at the same time. The ratio of the amplitude of the harmonic to that of the fundamental is

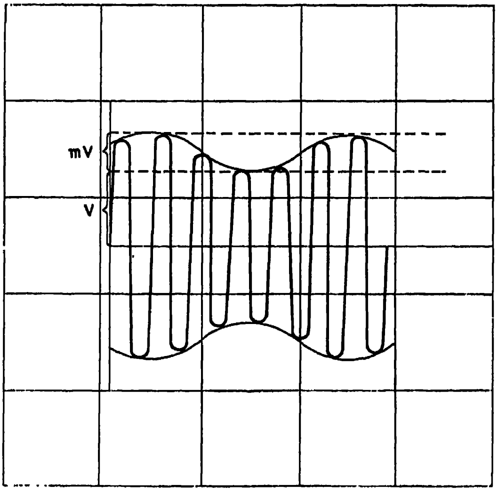


FIG. 10–8

$$\left(\frac{bV_m^2m^2}{4}\right)/bV_m^2m = m/4$$

and thus depends on the depth of modulation.

An extremely important application of the foregoing phenomenon is, of course, what is termed anode-bend detection in wireless telegraphy.

COMBINATION FREQUENCIES

By means of the rectifying properties of a triode it is possible to produce combination frequencies from two different high-frequency oscillations. Let the two frequencies be represented by $V_{m1} \sin p_1 t$ and $V_{m2} \sin p_2 t$. Then the resultant e.m.f. on the grid is

$$(V_{m1} \sin p_1 t + V_{m2} \sin p_2 t)$$

Substituting this in (10–17)

$$\begin{aligned} i_a &= i_Q + a(V_{m1} \sin p_1 t + V_{m2} \sin p_2 t) + b(V_{m1} \sin p_1 t + V_{m2} \sin p_2 t)^2 \\ &= i_Q + a(V_{m1} \sin p_1 t + V_{m2} \sin p_2 t) + b\left[\frac{V_{m1}{}^2}{2}(1 - \cos 2p_1 t)\right. \\ &\qquad \left. + 2V_{m1}V_{m2} \sin p_1 t \sin p_2 t + \frac{V_m{}^2}{2}(1 - \cos 2p_2 t)\right] \end{aligned}$$

or

$$\begin{aligned} & i_Q + \frac{b}{2}(V_{m1}{}^2 + V_{m2}{}^2) + a(V_{m1} \sin p_1 t + V_{m2} \sin p_2 t) \\ & \quad - \frac{b}{2}(V_{m1}{}^2 \cos 2p_1 t + V_{m2}{}^2 \cos p_2 t) \\ & \quad + bV_{m1}V_{m2} \cos (p_1 - p_2)t - bV_{m1}V_{m2} \cos (p_1 + p_2)t \qquad (10\text{–}19) \end{aligned}$$

With telephones in the anode circuit there will be no response to the high-frequency components and hence the current affecting the phones is

$$i_a = i_Q + \frac{b}{2}(V_{m1}{}^2 + V_{m2}{}^2) + bV_{m1}V_{m2} \cos (p_1 - p_2)t$$

Thus, because of the rectifying properties of the triode, a combination frequency $(p_1 - p_2)/2\pi$ is produced.

CUMULATIVE GRID RECTIFICATION

As previously mentioned, because the grid voltage/grid current characteristic of the triode is non-linear it is possible to produce another form of rectification in addition to that already treated. In order to effect this the arrangement of Fig. 10–9 is employed, where it will be noted that a capacitance and resistance are connected to the grid. As the rectifying process depends on the flow of grid current the point B is made positive to the cathode as shown. If Fig. 10–10 represents the grid current/grid voltage curve, then the initial potential of the grid due to R may be found in the following manner. If v_g is the grid potential, then the potential drop on R is $(v_b - v_g)$ and R is given by $(v_b - v_g)/i_g$ where i_g is the grid current. Hence

$$R = (v_b - v_g)/i_g = \cot \theta$$

and, if a straight line is drawn as shown, making an angle θ with the axis, this line, at its point of intersection with the grid current/grid voltage characteristic, will indicate the steady potential and current of the grid in the absence of an applied e.m.f. at AB.

We shall now suppose that a small high-frequency voltage is applied to AB, and, to simplify conditions, take P as the origin of co-ordinates. It is assumed that the values of R and C are so chosen that practically the whole of the applied e.m.f. is developed between grid and cathode of the valve. This e.m.f. will, in a similar manner to anode bend rectification, cause the increase in current above P to be greater than that below P, i.e. rectification will occur. The resulting d.c. component of the current will flow through R while the high-frequency components will pass through C. A result of the d.c. component is to increase the volt drop on R and decrease the mean grid potential. If v_0 is this decrease, and i_0 the d.c. component of the current, then $v_0 = Ri_0$, and the ultimate e.m.f. on the grid, as a result of applying an initial e.m.f. V_m sin pt, is $(V_m \sin pt - Ri_0)$. Hence

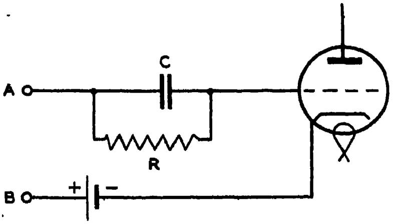


FIG. 10–9

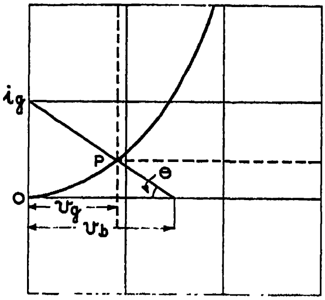


FIG. 10–10

$$i_g = a(V_m \sin pt - Ri_0) + b(V_m \sin pt - Ri_0)^2$$

remembering that P is taken as the origin. Now, i_g contains the d.c. component i_0, so equating d.c. components we have

$$i_0 = -aRi_0 + bR^2i_0^2 + \tfrac{1}{2}bV_m^2$$

or
$$bR^2i_0^2 - (1 + aR)i_0 + \tfrac{1}{2}bV_m^2 = 0$$

Solving, $i_0 = [(1 + aR) \pm \sqrt{(1 + aR)^2 - 2b^2R^2V_m^2}]/2bR^2$

Expanding the term under the radical by means of the binomial theorem, and, since V is small, neglecting all terms beyond the first two,

$$i_0 = \{(1 + aR) \pm (1 + aR)[1 - b^2R^2V_m^2/(1 + aR)^2]\}/2bR^2$$

As $i_0 = 0$ when $V_m = 0$, the positive sign is inadmissible and, therefore,

$$i_0 = \frac{\frac{1}{2}bV_m^2}{1 + aR}$$

Now as

$$i_g = av_g + bv_g^2$$

$$\frac{di_g}{dv_g} = a + 2bv_g$$

and if v_g is very small

$$\frac{di_g}{dv_g} = a = \frac{1}{R_s}$$

where R_s is the slope resistance of the grid-cathode circuit. Hence

$$i_0 = \frac{\frac{1}{2}R_s bV_m^2}{R_s + R} \quad . \quad . \quad . \quad (10\text{–}20)$$

and from this result we may consider the rectified current, i_0, to be due to a hypothetic generator of e.m.f. $\frac{1}{2}R_s bV_m^2$ and internal resistance R_s working into an external load resistance R.

If now a modulated e.m.f., represented by $V_m(1 + m \sin wt) \sin pt$, is applied to AB, the rectified current will contain low-frequency components as well as a direct component. Hence in this case the rectified current may be regarded as caused by a series of hypothetic generators, each of which is responsible for one component of the current. Each generator will possess an internal resistance R_s and deliver current to an external resistance R, assuming, of course, that C offers a relatively high impedance to low-frequency components. Now if we assume for the moment that R is zero, the total grid current is

$$i_g = aV_m(1 + m \sin wt) \sin pt + bV_m^2(1 + m \sin wt)^2 \sin^2 pt$$

or

$$aV_m(1 + m \sin wt) \sin pt + \tfrac{1}{2}bV_m^2(1 - \cos 2pt) \left[1 + 2m \sin wt + \frac{m^2}{2}(1 - \cos 2wt)\right]$$

If it is desired to record the modulating and direct components only, we see that the components concerned are those which were listed from (1) to (3) in dealing with anode bend rectification, except that in this case they are grid currents and not anode currents. Taking the low-frequency component $bV_m^2 m \sin wt$, this is the current which results when $R = 0$. Regarding (10–20), the current in this case for $R = 0$ is $\frac{1}{2}bV_m^2$. Hence we see that to obtain the current for $R \neq 0$ the current at $R = 0$ must be multiplied by $R_s/(R_s + R)$. Thus the low-frequency component of frequency $w/2\pi$ is

$$\frac{R_s b V_m^2 m \sin wt}{R_s + R}$$

and the low-frequency e.m.f. developed between grid and cathode is

$$\frac{RR_s b V_m^2 m \sin wt}{R_s + R} \quad . \quad . \quad . \quad (10\text{–}21)$$

Similarly the decrease in grid potential is

$$\frac{\frac{1}{2}RR_s b V_m^2(1 + m^2/2)}{R_s + R}$$

the second harmonic potential being

$$\frac{\frac{1}{4}RR_s b V_m^2 m^2 \sin 2wt}{R_s + R}$$

An extremely important application of the foregoing form of rectification is grid detection in wireless telegraphy. For this purpose the function of C is to provide a negligible impedance for the high-frequency components, so that practically the entire input voltage is applied between grid and cathode. The purpose of R is evident, for without it no low-frequency component can be developed at the grid. The amplitude of (10–21) increases with R and has a maximum value of $R_s b V_m^2 m$ when R is infinitely large. However, R must be definitely limited because, due to the shunting effect of C, amplitude distortion will occur if R is too large, low frequencies producing a higher e.m.f. at the grid than high frequencies. In practice the values of C and R should not exceed 0·0001 μF and 0·25 megohm respectively.

So far only the effects occurring in the grid circuit have been considered. As already shown, an e.m.f. v_g in the grid circuit is equivalent to an e.m.f. μv_g in the anode circuit. Thus various currents will flow in the anode circuit corresponding to those in the grid circuit and, if a suitable anode impedance is employed, amplified copies of the grid-cathode e.m.f.s will result.

Multi-electrode Valves

The evident advantages due to the introduction of a third electrode into the vacuum diode naturally led to valves with further electrodes. Such valves are denoted by characteristic terms such as tetrodes, pentodes, pentagrids, octodes, etc., according to the number of electrodes they possess. The earliest of the multi-electrode tubes was the tetrode which, in addition to the usual cathode and anode, has two grids. One of these is the normal control grid, while the other is generally held at a fixed positive potential relatively to the cathode. Typical circuit connexions are shown by Fig. 10–11.

A property of primary importance possessed by the tetrode, or "screen-grid," valve is a very small inter-electrode capacitance between the control grid and anode. The screen grid is similar in structure to the control grid and is placed outside the latter, i.e. between the control grid and anode. The electrostatic screening of the second grid is such that it may reduce the grid-anode capacity by a factor of 1000 or even more, although the meshes of the screen are sufficiently wide as not to offer mechanical impedance to the electrons. As an example it may be stated that the grid-anode capacitance of the Ferranti VPT4 valve is only 0·002 $\mu\mu$F, whereas this capacitance for an ordinary triode might be as high as 10 $\mu\mu$F.

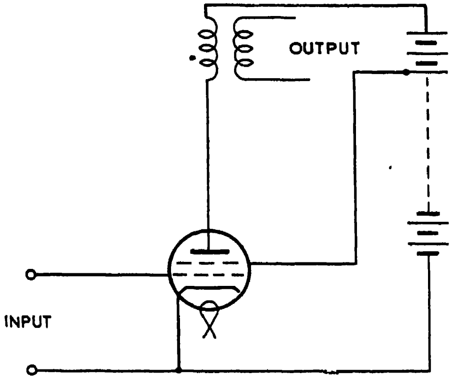


Fig. 10–11

Because of the screening action of the second grid, the electric field at the cathode resulting from the anode potential is extremely small. The result is that the anode has but little effect on the space-charge current drawn from the cathode, control of this being almost entirely effected by the control grid. Thus, in contrast with a triode, the anode of a tetrode merely collects the current passing through the screen, whereas in the former case the anode both controls and collects the current.

As the anode current is but slightly affected by a change in anode potential, it follows from the definition of slope resistance that this quantity for a tetrode must be relatively high. Thus, in the case of the VPT4, the slope resistance is approximately 1·0 megohm. Although the inter-electrode capacitance and slope resistance of the tetrode differ so greatly from those of a triode, the mutual conductances of both valves are similar. This is because the disposition of the cathode and control grid is similar in both cases. As $\mu = \rho g_m$ it follows that the tetrode possesses a high amplification factor and thus is a much better amplifier than a triode. For the VPT4 valve $g_m = 2{\cdot}0$ mA per volt and thus $\mu = 2000$.

TETRODE CHARACTERISTICS

As stated above, the mutual characteristics of triodes and tetrodes are similar. However, if a family of mutual characteristics for a tetrode is plotted, it will be found that curves for different anode voltages lie very close together. This, of course, is due to the relatively small influence of the anode voltage on the anode current. Although the mutual characteristics of triodes and tetrodes are similar, the anode characteristics are very different. In considering these characteristics it is assumed that the screen voltage is maintained constant at a value equal to about 75 per cent of the normal anode voltage. With a given voltage on the control grid, as the anode voltage is raised from zero the anode current continually increases until, at about 10–20 volts, it reaches a maximum and then commences to fall. The anode current continues to decrease as the anode voltage is increased until the latter approaches a value of the same magnitude as that of the screen. At this point a minimum occurs and thereafter the current increases with voltage until saturation sets in. A set of typical anode characteristics is shown by Fig. 10–12, from which it is evident that over a considerable portion of the curves the valve possesses a negative slope resistance.

The cause of the negative slope is secondary emission from the anode due to the impact of primary electrons. Thus, when the anode voltage is higher than about 10 volts, the electrons which strike the anode release an appreciable number of secondary electrons from its surface. Because the screen potential is higher than that of the anode the secondaries are attracted to the screen, resulting in an increase in screen current and a decrease in anode current. The negative slope of the anode characteristics depends on the number of secondary electrons released per primary electron and increases in steepness with an increase in this ratio. Under some circumstances the anode current may become negative. As the anode voltage is increased and becomes of the same order as that of the screen, the secondary electrons formed at the anode are immediately drawn back to it, and from this point, for higher anode voltages, play a decreasing part in the behaviour of the valve.

Owing to the shielding effect of the screen, the field due to the anode has little effect on the cathode space-charge current, with the result that the current from the cathode is practically constant over the range of variation of the anode voltage. Hence the sum of the screen and anode currents tends to be constant as shown by Fig. 10–12. It will be noted that the total current tends to rise slightly as the anode voltage increases. This is because secondary electrons are liberated from the screen and are collected by the anode.

For normal purposes tetrodes must be operated with the anode

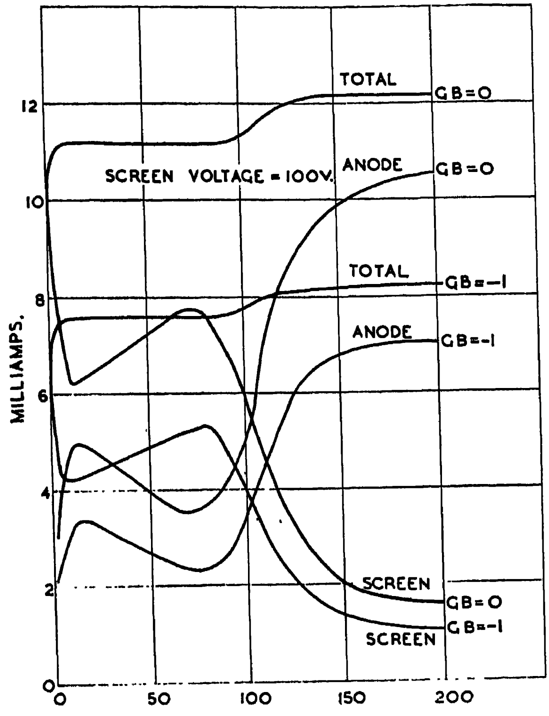


Fig. 10–12

potential in excess of that of the screen. However, operation of the valve where the slope resistance is negative has several important applications, one of which is described on page 395. Operated in this region the valve is capable of supplying power and may be employed to produce sustained oscillations. Applied in this manner, the valve is termed a *dynatron*. It may be mentioned that a negative slope resistance is also possessed by a triode when the grid is operated at a higher potential than the anode.

TETRODE INPUT IMPEDANCE

If the ideal assumption is made that the tetrode grid-anode capacitance is zero, then, from (10–14), it will be seen that $\mathbf{Z}_i$ is infinite. However, the capacitance formed between the two grids is in parallel with C_1 and hence the input capacitance is

$$C_1 + C_4$$

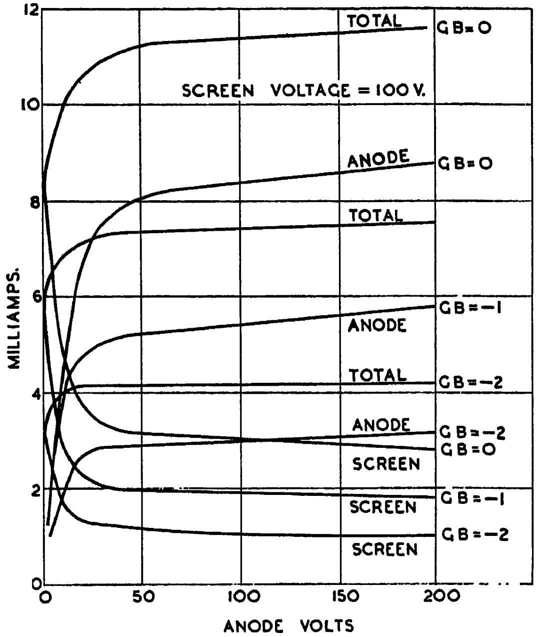


FIG. 10–13

where C_4 is the capacitance between the grids. With the above assumption it will be appreciated that spontaneous oscillation cannot occur.

THE PENTODE

In spite of the several advantages of the tetrode it possesses the disadvantage that for the usual applications the anode voltage must not be allowed to decrease to the extent that it approaches the screen voltage. Thus, when employed as an amplifier, the anode voltage swing must be distinctly limited in order that the kinks in the anode characteristics and consequent distortion shall be avoided.

The foregoing disadvantage is overcome by the introduction of a further grid situated between the screen and anode. This fifth electrode is termed a *suppressor* grid, because it suppresses the secondary emission. The suppressor is usually connected either internally or externally to the cathode and hence is at the potential of the latter. Thus, because of the direction of the field between suppressor and anode, secondaries produced at the screen are also returned to this. However, the electrons from the cathode are not appreciably affected and readily pass through both screen and suppressor to the anode.

A typical set of anode characteristics for a pentode is shown by Fig. 10–13, the valve being identical with that of Fig. 10–12, except for the inclusion of the suppressor grid. It will be noted that the kinks are absent and also that the screen current tends to remain constant except in the vicinity of the origin. The magnitude of the screen current is determined by the number of electrons intercepted by the screen wires. As with the tetrode, the anode current is but slightly affected by variations in anode voltage and hence the pentode has a relatively large slope resistance and amplification factor. The exact values of these quantities depend on the purpose for which the pentode is intended, for it may be employed as either a voltage amplifier or a power valve. In the former case the values are similar to those already quoted on page 329 for the tetrode. As explained on page 367, the power sensitivity of a pentode is much higher than that of a triode and for this reason the pentode is frequently employed as a power valve. In these circumstances the slope resistance is of the order of 10^5 ohms, the amplification factor being of the order of 100. The value of the mutual conductance is similar to that of a triode.

VARIABLE-μ VALVES

By modifying the normal structure of the control grid it is possible to provide a valve with a variable amplification factor. The modification is usually applied to either tetrodes or pentodes, and takes the form of having either the grid-cathode spacing, the spacing between the grid wires, or the diameter of the grid wires non-uniform along the length of the grid structure. By this means various parts of the grid will exercise different degrees of control over the anode current. Thus, one part of the grid will cut off the anode current at a different grid voltage than some other part, the former part, for example, having more turns per centimetre than the latter part. The result of this is that the anode current will decrease more and more slowly as the grid becomes more negative. Thus the mutual characteristics of a valve constructed in the above manner show an asymptotic approach to the axis as indicated by Fig. 10–14. Different parts of the grid may be considered as conferring on the valve different amplification factors, and for this reason such valves are termed variable-μ valves. The principal application of these valves is for volume control in radio-receiving circuits. The steady grid bias is varied in order to vary the mutual conductance, and hence the amplification factor and output.

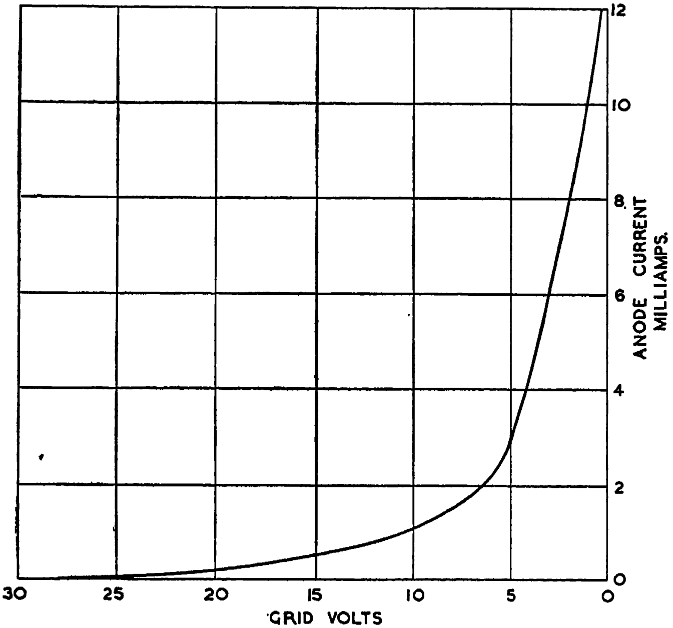


FIG. 10–14

OTHER MULTI-ELECTRODE VALVES

In addition to diodes, triodes, tetrodes, and pentodes, other valves, exist which, generally, are combinations of the foregoing types in one envelope. Thus the class B valve (dealt with in Chapter XI) consists of two triodes.

For frequency conversion in superheterodyne receivers a single-valve frequency changer is often employed, one form consisting of a triode-pentode as shown by Fig. 10–15. The triode constitutes a local oscillator, while the input signal is applied to the pentode. Both triode and pentode have a common cathode. From Fig. 10–15 it will be noted that the voltage between the pentode control grid and cathode consists of the input signal voltage plus that of the grid coil voltage of the oscillator. For this reason the pentode is sometimes termed a "mixer." The result of these two voltages on the pentode anode current is given by (10–19), from which it will be noted that various frequency combinations result. That of importance is the frequency difference of the two voltages and to this resultant frequency (termed the intermediate frequency) the anode circuit of the pentode is tuned. The other components are by-passed to the cathode by the decoupling condenser C. As $(p_1 - p_2)/2\pi$ is due to the rectifying properties of the pentode this valve is often referred to as the *first detector* when employed in a superheterodyne circuit. The method of combining the two frequencies in Fig. 10–15, i.e. that of the signal and local oscillator, is termed *cathode injection*.

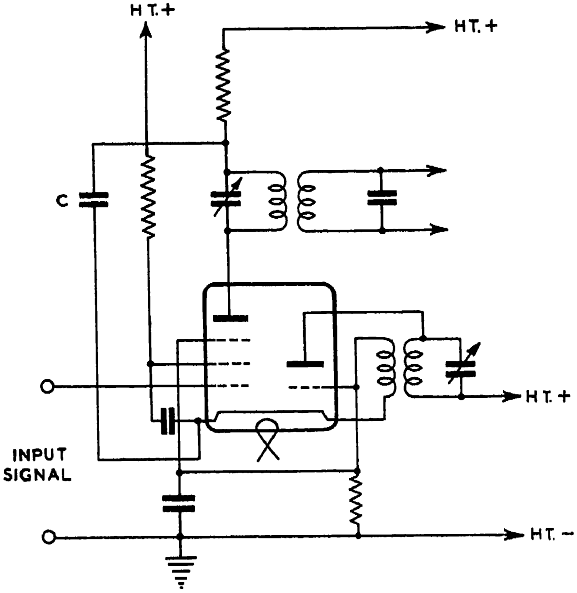


Fig. 10–15

THE OCTODE FREQUENCY-CHANGER

The signal and oscillator voltages of Fig. 10–15 are directly coupled. By means of an octode or heptode valve these voltages may be electron-coupled. An octode is shown by Fig. 10–16 and will be seen to consist of a cathode and anode with no less than six grids. The structural arrangement is such that the grids and anode are concentric with the cathode, the anode being the outside electrode. Considering Fig. 10–16, the cathode and grids 1 and 2 form a triode employed for generating the local oscillator voltage. Because of the positive potential on grid 3 and the grid-like structure of the triode anode, i.e. grid 2, the electrons pass into the space between grids 3 and 4. However, due to a negative bias applied to the control grid 4, the electrons are repelled, with the result that the region between grids 3 and 4 is occupied by a pulsating space-charge of heterodyne frequency. This space-charge may be regarded as a virtual cathode from the viewpoint of the anode, the latter drawing its electrons from this cathode. In the absence of a signal voltage the electron stream to the anode will oscillate at heterodyne frequency. However, when a signal is impressed on the control grid 4, the signal frequency and that of the local oscillator will be mixed

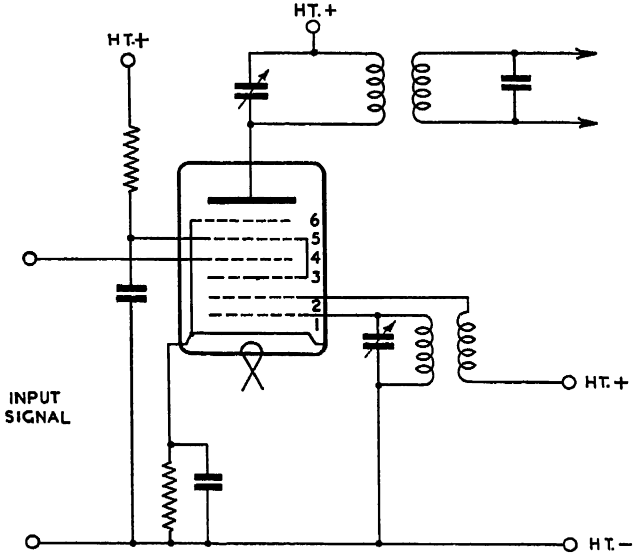


FIG. 10–16

by electron coupling and combination frequencies produced. The anode circuit is tuned to the frequency difference, the various other components being by-passed to earth by the decoupling condenser. The virtual cathode with grids 4, 5, 6, and the anode form a pentode mixer or first detector and this may be designed to have variable-μ characteristics if desired.

The heptode is similar in construction and principle to the octode, but has no suppressor grid.

BIBLIOGRAPHY

The Admiralty Handbook of Wireless Telegraphy (H.M.S.O.).
The Superheterodyne Receiver, A. T. Witts (Pitman).
Thermionic Valves in Modern Radio Receivers, A. T. Witts (Pitman).

CHAPTER XI

VALVE AMPLIFIERS

Voltage Amplification Factor

CONSIDERING the circuit of Fig. 10–6, if a voltage V_g is applied between grid and cathode the resulting current in the anode circuit produces a voltage across $\mathbf{Z}_a$ equal to $\mathbf{Z}_a I_2$. Consequently the ratio $\mathbf{Z}_a I_2 / V_g$ is termed the *voltage amplification factor*, and to determine this quantity we must first determine I_2. From (10–12) and (10–13)

$$I_2 = \frac{V_g + \dfrac{j(1 + \mu + \rho/\mathbf{Z}_a)}{pC_2[\rho - j/pC_2 \,.\, (1 + \rho/\mathbf{Z}_a)]} V_g}{\mathbf{Z}_a}$$

Hence, denoting the voltage amplification factor by $\mathbf{M}$,

$$\mathbf{M} = 1 + \frac{j}{pC_2} \frac{(1 + \mu + \rho/\mathbf{Z}_a)}{[\rho - j/pC_2 \,.\, (1 + \rho/\mathbf{Z}_a)]}$$

which for low frequencies reduces to

$$-\frac{\mu \mathbf{Z}_a}{\mathbf{Z}_a + \rho} \quad . \quad . \quad . \quad . \qquad (11\text{–}1)$$

From this result it is at once apparent that $\mathbf{M}$ must always be less than μ, the two only being equal if $\mathbf{Z}_a = \infty$. In the event of $\mathbf{Z}_a$ being a pure resistance, then

$$M = \mu \frac{R_a}{R_a + \rho}$$

if we disregard the sign. In this case it is evident that if a high amplification is required, R_a should be large. However, as R_a is increased, the anode potential is decreased if the high-tension supply voltage is maintained constant. The result of this is that, in practice, M seldom exceeds $0{\cdot}8\mu$.

In order to discuss amplification to some useful purpose it is necessary to take into consideration the effect of any circuit which may be connected either across $\mathbf{Z}_a$ or between the cathode and anode of the valve. For example, the amplification obtainable from a single valve is frequently inadequate for the purpose in view, and in such cases the amplification may be greatly increased by adding further valves, the output voltage of one constituting the input voltage of the next. If M_1, M_2, M_3, etc., are the voltage-amplification factors of succeeding valves (M_1 in this case being termed the stage gain), then the overall gain is $M_1M_2M_3$ which, if $M_1 = M_2 = M_3$, is equal to $M_1{}^3$. Moreover, the promise held out in this manner of enormous gains is seldom realized in practice, because beyond two or three stages instability arises which definitely limits the amplification which may be obtained.

Resistance-capacitance Coupled Amplifiers

With this form of amplification $\mathbf{Z}_a$ takes the form of a "pure" resistance and the valve is coupled to the output circuit (or input circuit of the succeeding valve) by means of a coupling condenser C, as shown by Fig. 11-1. The purpose of C is to isolate the output circuit from the d.c. component of the anode potential while allowing the a.c. component a free path. If the output circuit is the grid and cathode of another valve, then the arrangement of Fig. 11-2 is employed where R_1 is a high resistance which permits the grid to be maintained at an appropriate mean d.c. potential. This potential is negative and of such value as to prevent the grid from becoming positive when an alternating e.m.f. is impressed on it from the preceding valve. R_1 also serves the further purpose of returning any charge acquired by the grid to the cathode, thus preventing the grid

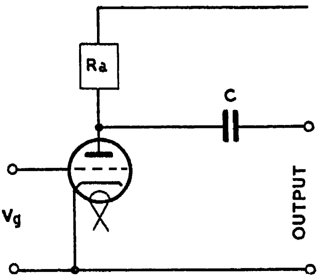


Fig. 11-1

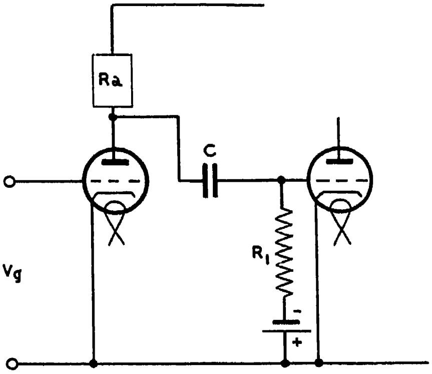


Fig. 11-2

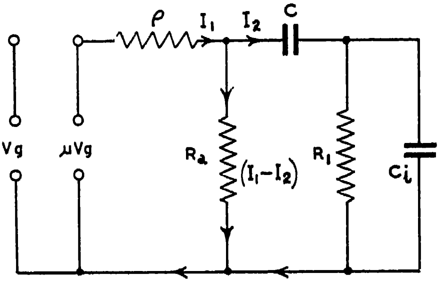


Fig. 11-3

from floating. For this reason R_1 is sometimes referred to as a "grid leak."

The equivalent circuit of Fig. 11–2 is shown by Fig. 11–3 where C_i is the input capacitance of the succeeding valve, C_i being composed of C_1 in parallel with a circuit whose impedance is given by (10–14). Now at audio-frequencies the shunting effect of C_i is negligible, and in this case it may be neglected. Hence, for audio-frequencies the circuit equations are

$$\rho I_1 + R_a(I_1 - I_2) = \mu V_g \quad . \quad . \quad . \quad (11\text{–}2)$$

$$-jI_2/pC + R_1 I_2 = R_a(I_1 - I_2) \quad . \quad . \quad (11\text{–}3)$$

The value of M being given by $R_1 I_2/V_g$, it is necessary to determine I_2. From (11–2) and (11–3) we have

$$I_2 = \frac{\mu V_g R_a}{(\rho + R_a)\left(R_1 + \dfrac{\rho R_a}{\rho + R_a} - \dfrac{j}{pC}\right)}$$

and

$$\mathbf{M} = \frac{\mu R_a R_1}{(\rho + R_a)\left(R_1 + \dfrac{\rho R_a}{\rho + R_a} - \dfrac{j}{pC}\right)}$$

Now, if C is sufficiently large, $\mathbf{M}$ will be independent of frequency, and in this case

$$M = \frac{\mu R_a R_1}{(\rho + R_a)\left(R_1 + \dfrac{\rho R_a}{\rho + R_a}\right)}$$

$$= \frac{\mu \dfrac{R_a R_1}{R_a + R_1}}{\rho + \dfrac{R_a R_1}{R_a + R_1}}$$

$$= \frac{\mu R_{a1}}{\rho + R_{a1}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11\text{–}4)$$

where R_{a1} is the resistance of R_a and R_1 in parallel.

At all but extremely low frequencies the effect of C is negligible, but at high frequencies the shunting effect of C_i must be taken into account. The equivalent circuit is shown by Fig. 11–4 and the circuit equations of this are

$$\rho I_1 + R_{a1}(I_1 - I_2) = \mu V_g$$

$$-jI_2/pC_i = R_{a1}(I_1 - I_2)$$

from which
$$I_2 = \frac{\mu V_g R_{a1}}{(\rho + R_{a1})\left(\frac{\rho R_{a1}}{\rho + R_{a1}} - \frac{j}{pC_i}\right)}$$

In this case
$$\mathbf{M} = -\frac{jI_2}{pC_i}/V_g$$
$$= \frac{-j\mu R_{a1}}{pC_i(\rho + R_{a1})\left(\frac{\rho R_{a1}}{\rho + R_{a1}} - \frac{j}{pC_i}\right)}$$

which leads to
$$\mathbf{M} = \frac{\mu}{j\rho pC_i + (\rho + R_{a1})/R_{a1}}$$

and
$$M = \frac{\mu}{\sqrt{\rho^2 p^2 C_i^2 + (\rho + R_{a1})^2/R_{a1}^2}} \quad . \quad . \quad (11\text{–}5)$$

In practice, R_{a1} is usually large compared with ρ, so that $(\rho + R_{a1})/R_1$ will be approximately equal to unity. Hence, if the amplification is to be independent of frequency, $\rho^2 p^2 C_i^2$ must be small compared with unity. Practical values of ρ and C_i are such that, generally speaking, resistance-capacitance coupling is unsuitable for frequencies much above 5×10^4 cycles per second. Hence it finds its chief applications in the audio-frequency range, where, providing C is not too small, a voltage-amplification factor almost independent of frequency may be obtained.

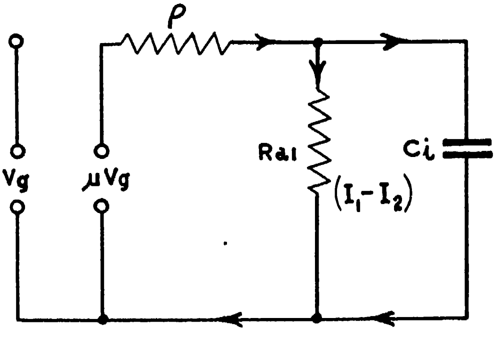


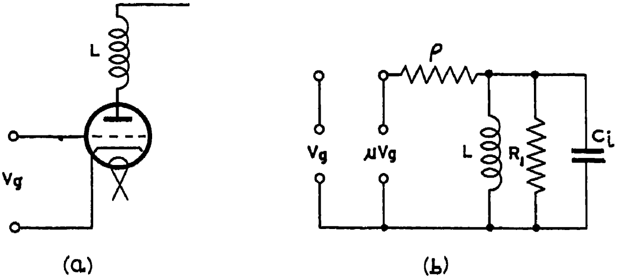
FIG. 11-4

VALUES OF R_a AND R_1

If $\rho^2 p^2 C_i^2$ is small compared with $(\rho + R_{a1})^2/R_{a1}^2$ it will be noted that (11–5) approximates to (11–4). Hence, in order that the amplification shall be high, it is necessary that both R_a and R_1 are large compared with ρ. However, neither of these quantities may be increased indefinitely, for as we have already seen, increasing R_a lowers the valve anode potential. If R_a is limited, then as R_1 is in parallel with R_a, there is no point in having R_1 greatly in excess of R_a. In practice, R_a is usually limited to about 10^5 ohms and R_1 to 1 megohm.

Choke-capacitance Coupling

In order to avoid the d.c. drop in voltage which occurs on R_a, it is possible to replace this resistance with a choke. If high-frequency amplification is desired the choke must be air-cored, while an iron-cored choke is employed for low-frequency work. A circuit employing a choke is shown by Fig. 11-5 (*a*), the equivalent circuit being given by (*b*). In the latter, C_i includes the self-capacitance of the choke and also the input capacitance of the next valve. In this case we shall take $\mathbf{Z}_a$ to be the impedance of L, C_i, and R_1 in parallel. Thus



(a) (b)

FIG. 11-5

$$\frac{1}{\mathbf{Z}_a} = \frac{1}{R_1} + \frac{1}{jpL} + jpC_i = \frac{1}{R_1} + j(pC_i - 1/pL)$$

Now
$$\mathbf{M} = \frac{\mu \mathbf{Z}_a}{\mathbf{Z}_a + \rho} = \frac{\mu}{1 + \rho/\mathbf{Z}_a}$$

and, substituting, we obtain

$$\mathbf{M} = \frac{\mu}{1 + \rho/R_1 + j\rho(pC_i - 1/pL)}$$

From this result it is evident that $\mathbf{M}$ will be a maximum when conditions are such that $pC_i = 1/pL$, i.e. when $p^2 = 1/C_iL$. Normally, of course, the values of C_i and L are fixed, and as the amplifier will, generally, be required to operate over a wide range of frequencies it is desirable that the quantity $(pC_i - 1/pL)$ shall be as small as possible. This quantity may be written $(p^2C_iL - 1)/pL$ and, as C_iL is constant, it follows that the ratio C_i/L should be as small as possible.

As stated above, for low-frequency work iron-cored chokes must be employed because of the low value of p. As the direct current flows through the choke, this must be so designed that magnetic saturation does not occur.

Tuned Anode Coupling

As an alternative to having C_i and L fixed, it is possible to have C_i variable by connecting a variable capacitance across L and tuning the circuit C_iL to resonance. Now it is well-known that a parallel circuit of this type at resonance behaves as a pure resistance L/C_iR_L, where R_L is the ohmic resistance of the choke. Hence, in this case we have

$$M = \frac{\mu L/C_iR_L}{L/C_iR_L + \rho}$$

Transformer-coupled Amplifiers

By employing transformer coupling in an amplifier at least two very important advantages are gained: (1) a voltage-amplification factor higher than that of the valve may be obtained, (2) it is possible to dispense with C, the coupling condenser, because the primary and secondary windings are isolated from each other. Before discussing transformer coupling in detail, it is first desirable to consider some properties of transformers.

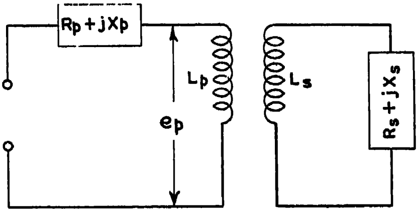


FIG. 11-6

Considering the transformer circuit shown by Fig. 11-6, let impedances be connected in both primary and secondary windings, as shown. In the case of the primary winding, R_p is the primary resistance and X_p the primary reactance. Actually, X_p is due to the leakage inductance of the primary winding. In the case of the secondary, R_s is the sum of the secondary winding and load resistances, while X_s includes the effects of the secondary leakage inductance and self-capacitance as well as the load reactance. Now, if perfect coupling exists between the primary and secondary inductances L_p and L_s, then

$$m^2 = L_p L_s$$

where m is the mutual inductance. Also, if L_p and L_s are assumed to be respectively proportional to the squares of the numbers of turns on the primary and secondary windings, then $n = \sqrt{L_s/L_p}$, where n is the transformer turns ratio. If e_p is the voltage across L_p, then

$$jpL_pI_p + jpmI_s = e_p$$

$$jpL_sI_s + jX_sI_s + R_sI_s + jpmI_p = 0$$

Solving for I_p and I_s, we obtain

$$I_p = e_p\left[\frac{1}{jpL_p} + \frac{1}{(R_s + jX_s)/n^2}\right]$$

$$I_s = \frac{ne_p}{R_s + jX_s}$$

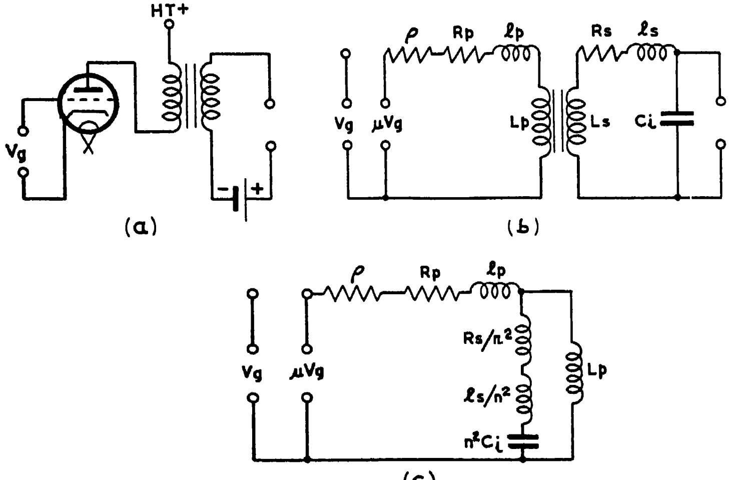


FIG. 11-7

From this result it is evident that from the viewpoint of the primary the circuit consists of the impedance $(R_p + jX_p)$ in series with the combination of jpL_p and $(R_s + jX_s)/n^2$ in parallel. Also, it is clear that the voltage across L_s is ne_p.

The usual method of coupling two valves by means of a transformer for audio-frequency working is shown by Fig. 11-7 (a), the equivalent circuit being shown by (*b*). In (*b*) C_i includes the primary and secondary self-capacitance, the inter-winding capacitance, and the input capacitance of the next stage. l_p and l_s are respectively the primary and secondary leakage inductances. The circuit of (*b*) can be modified still further to the equivalent circuit of (*c*). Referring to (*c*), at low frequencies the branch containing n^2C_i will behave as a high-capacitive reactance, so that its shunting effect on L_p may be neglected. For example, let $L_p = 50$ henrys, $l_s = 4{\cdot}5$ henrys, $C_i = 100\ \mu\mu\text{F}$, and $n = 3$. Then at 100 cycles per second the reactance of L_p is 31,400 ohms, and that of n^2C_i 1·77 megohms. Hence, as the impedance of R_s/n^2 and l_s/n^2 is negligible compared with that of n^2C_i, the voltage on the latter will be practically identical with that on L_p. Thus the input voltage to the next stage is approximately

$$\frac{jpL_pn\mu V_g}{\rho + jpL_p}$$

and

$$\mathbf{M} = \frac{jpL_pn\mu}{\rho + jpL_p} \quad . \quad . \quad . \quad . \quad (11\text{–}6)$$

where ρ may be considered to contain R_p. From this result it is evident that at very low frequencies **M** will be small, falling to zero as p approaches zero.

Now as the frequency increases, the reactance of L_p will increase, while that of the branch containing n^2C_i will decrease. However, even at 1000 cycles per second the voltage on n^2C_i is still practically identical with that on L_p and **M** will, therefore, still be given by (11–6) if the effect of l_p is neglected. It may be noted that when

$$\frac{1}{pn^2C_i} = pL_p$$

a condition of resonance exists, L_p and its shunt then behaving as a pure resistance equal to

$$\frac{L_p}{C_iR_s}$$

For the values given, this resonant effect occurs at about 715 cycles per second, but is practically without effect on **M**.

As the frequency increases above 1000 cycles per second the reactance of L_p will tend to be such that it may be neglected in determining **M**. In this case we may include R_s/n^2 in ρ and write

$$\mathbf{M} = \frac{-jn\mu/pn^2C_i}{\rho + (jpl - j/pn^2C_i)}$$

where $l = l_p + l_s/n^2$. The magnitude of **M** is

$$M = \frac{n\mu/pn^2C_i}{\sqrt{\rho^2 + (pl - 1/pn^2C_i)^2}}$$

or

$$= \frac{n\mu X_c}{\sqrt{\rho^2 + (X_l - X_c)^2}}$$

Differentiating with respect to p and equating to zero for a maximum, it is found that the latter occurs for

$$X_c = \frac{2X_l^2 + \rho^2}{2X_l}$$

from which

$$p = \frac{\sqrt{2l/n^2C_i - \rho^2}}{\sqrt{2}l}$$

and as ρ^2 is usually negligible compared with $2l/n^2C_i$

$$p = \frac{1}{\sqrt{ln^2C_i}}$$

and

$$f = \frac{1}{2\pi}\cdot\frac{1}{\sqrt{ln^2C_i}} \quad . \quad . \quad . \quad (11\text{–}7)$$

which gives the resonant frequency f. For the values of l and C_i given above, $f = 5340$ cycles per second. At this frequency we have

$$M_{max} = \frac{n\mu}{pn^2C_i\rho} = \frac{n\mu}{\rho}\sqrt{\frac{l}{n^2C_i}} \quad . \quad . \quad (11\text{–}8)$$

which, for the values given, is $2{\cdot}22\,n\mu$. Beyond the resonant frequency the value of M tends to fall rapidly and it is evident that this peak in the voltage-amplification curve may be a very undesirable feature.

From the foregoing analysis of transformer coupling as applied to the audio-frequency range several outstanding features engage the attention.

1. In order to avoid a decline in the amplification at low frequencies the transformer primary impedance should be high. Consideration of (11–6) shows that L_p should be such that pL_p is high compared with ρ.

2. In order that the voltage amplification shall be as uniform as possible over a given range of frequencies, it is desirable that the resonant frequency shall lie beyond the upper frequency limit; i.e. f, as given by (11–7), should be high. This may be effected by making l and C_i as small as possible. In order to obtain a low value of l the coupling of the transformer primary and secondary windings must be of a high order.

3. Referring to (11–8) it is evident that to avoid excessive peaking l again should be low. Furthermore, M_{max} varies inversely as ρ. Now in deriving (11–8), ρ was taken to contain R_p and R_s/n^2. For reasons to be given presently it is not desirable to increase the slope resistance of the valve and R_p to decrease M_{max}. Hence to effect this R_s may be made high, either by winding the secondary of high-resistance wire or, alternatively, by connecting a resistance in parallel with the secondary windings.

As the overall amplification is $n\mu$, it might appear that n and μ should be as high as possible. However, a high value of μ also means a high slope resistance and this will cause a falling off of the amplification at low frequencies. It is, of course, for this reason that R_p is not increased to reduce M_{max}. If n is large, this means either a small primary or a large secondary. With the first alternative L_p will be small, and with the second C_t will be large. Hence in practice n seldom exceeds 5, and is usually less.

One method of ensuring that l and C_t shall be small is to form the transformer core of high-permeability material such as Mumetal or Permalloy. If this is done then, for a given primary inductance and voltage ratio, the numbers of primary and secondary turns will be relatively small. This, of course, leads to small values for the leakage inductance and self-capacitance.

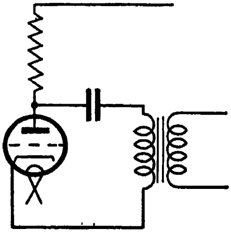


FIG. 11–8

From Fig. 11–7 (*a*) it will be noted that the d.c. component of the valve current traverses the transformer primary winding. An effect of this is to reduce the incremental permeability of the iron core and thus reduce the value of the primary inductance.* This may be avoided by the arrangement of Fig. 11–8, known as *parallel feed*. Here the function of the condenser is to isolate the transformer primary from the d.c. component. Its reactance should be such as to provide a negligible impedance to the lowest frequency a.c. component. Also the value of the anode-feed resistance should be high compared with the transformer primary reactance.

High-frequency Transformer Coupling

For amplification of frequencies above the audio range an iron-cored transformer is not suitable because of the iron losses which occur. The effect of such losses may be simulated by a resistance in parallel with the primary winding, the value of this resistance decreasing with an increase in frequency. This, of course, has a shunting effect on the primary, thus reducing the amplification. A further effect of eddy currents at high frequencies is to reduce the area of the core which is penetrated by the magnetic flux. This, of course, reduces the primary inductance. For these reasons air-cored transformers are usually employed for high-frequency amplification, or transformers possessing a core formed of iron dust bonded with a synthetic resin.

---

* "Iron-Cored Inductances," F. G. Spreadbury, *The Electrical Engineer*, Vol. VI., p. 768.

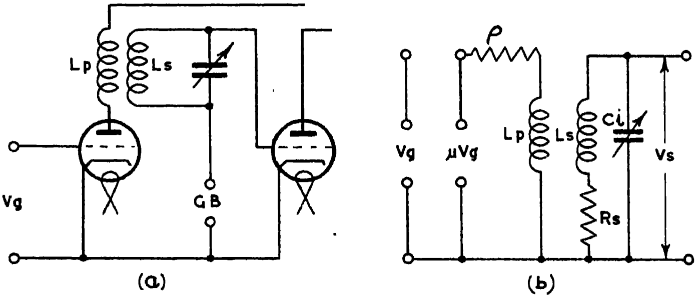


Fig. 11–9

Most of the analysis as previously applied to audio-frequency amplification holds good for frequencies outside the audio range. However, as in this case p is large, L_p, L_s, and C_i will be correspondingly small. Also, because of the absence of an iron core, or, at the best, the presence of one of low permeability, the coupling factor of the transformer will be much lower than that for an audio-frequency transformer. This, naturally, leads to a relatively high value of l.

In the case of the audio-frequency amplifier this is frequently employed for the simultaneous amplification of a wide variety of frequencies such as are found in speech or music. High-frequency amplifiers, however, often deal with only one frequency at a time and hence the possibility exists of tuning the amplifier to a particular frequency. In this case the amplifier often takes the form shown by Fig. 11–9 (*a*), the equivalent circuit being given by (*b*). Here C_i includes the self and interwinding capacitances, the input capacitance of the following valve, and a variable capacitance for tuning L_sC_i to resonance. In determining V_s, R_p will be neglected, as it is negligible compared with ρ. If m is the mutual inductance between the primary and secondary windings, then the circuit equations are

$$jpL_pI_p + jpmI_s + \rho I_p = \mu V_g \quad . \quad (11\text{–}9)$$

$$jpL_sI_s + jpmI_p + R_sI_s - jI_s/pC_i = 0 \quad . \quad (11\text{–}10)$$

Solving for I_s, we obtain

$$I_s = \frac{-jpm\mu V_g}{(\rho + jpL_p)\left[(R_s + jpL_s - j/pC_i) + \dfrac{m^2p^2}{\rho + jpL_p}\right]}$$

and $V_s = -jI_s/pC_i$

$$= \frac{pm\mu V_g}{pC_i(\rho + jpL_p)\left[(R_s + jpL_s - j/pC_i) + \dfrac{m^2p^2}{\rho + jpL_p}\right]}$$

taking the absolute value.

Now when the secondary circuit is tuned to resonance $(jpL_s - j/pC_i) = 0$. Therefore

$$V_s = \frac{m\mu V_g}{C_iR_s(\rho + jpL_p) + C_im^2p^2}$$

$$= \frac{m\mu V_g}{C_iR_s\rho + jpC_iL_pR_s + C_im^2p^2}$$

Multiplying numerator and denominator by L_s/m

$$V_s = \frac{L_s\mu V_g}{L_sC_iR_s\rho/m + jpL_sC_iL_pR_s/m + L_sC_imp^2}$$

But $L_sC_i = 1/p^2$ and

$$V_s = \frac{L_s\mu V_g}{L_sC_iR_s\rho/m + jL_pR_s/mp + m}$$

or, as L_pR_s/mp is generally extremely small,

$$V_s = \frac{L_s\mu V_g}{C_iR_s\rho/a + aL_s} \quad . \quad . \quad . \quad . \quad . \quad (11\text{–}11)$$

where $a = m/L_s$

$$\text{Now} \quad \frac{m}{L_s} = \frac{m}{\sqrt{L_pL_s}}\sqrt{\frac{L_p}{L_s}} = k/n$$

and hence $a = k/n$, where k and n are, respectively, the transformer coupling factor and transformer ratio. In order to find the maximum value of (11–11) we differentiate V_s with respect to a and equate to zero. Thus

$$a = \sqrt{C_iR_s\rho/L_s} \quad . \quad . \quad . \quad (11\text{–}12)$$

Now C_i varies inversely as the square of the frequency and, as a is presumed to be fixed by the transformer construction, V_s can only be a maximum for one particular frequency. Hence in practice it is customary to design the transformer for a frequency intermediate to the extreme values likely to be met.

Now C_iR_s/L_s is the reciprocal of the effective resistance which the circuit L_sC_i offers at resonance. Denoting this resistance by R_r

$$a = \sqrt{\rho/R_r}$$

and substituting in (11–11)

$$V_s = \tfrac{1}{2}\mu V_g\sqrt{R_r/\rho}$$

From this result it is evident that for high amplification to be obtained, R_r should be large. This, of course, requires that R_s shall be small. Also a valve with a high value of μ is desirable.

Distortion in Amplifiers

Various forms of distortion may occur in amplifiers, one form, i.e. frequency distortion, having already been dealt with. With this form the ratio of the amplitudes of the output and input voltages is constant at a given frequency, but varies with different frequencies. Of the various types of amplifiers discussed that employing resistance-capacitance coupling is the best where freedom from frequency distortion is desired. An extremely important form of distortion occurs when the amplitude of the output voltage, at a given frequency, is not proportional to the amplitude of the input voltage. This form is known as *amplitude*, *non-linear*, or *harmonic* distortion and is characterized by the introduction of harmonics into the output voltage wave which are not present in the input wave. The cause of this arises from the fact that the i_a/v_g characteristics of Fig. 10–4 are not strictly linear. As we have already seen, the relationship between i_a and v_g, with a pure resistance in the anode circuit, is given by

$$i_a = \frac{\mu v_g}{\rho + R_a} \quad . \quad . \quad . \quad (11\text{–}13)$$

and, if μ and ρ are constant, this relation is linear. The curve drawn from (11–13) is known as the *dynamic* characteristic and, because of variations in μ and ρ (particularly the latter), is non-linear. However, the variations in ρ are minimized by the presence of R_a and hence the dynamic characteristic departs less from linearity than does the static characteristic, the latter being given by $i_a = \mu v_g/\rho$.

For the reasons just given the dynamic characteristic is more accurately represented by

$$i_a = av_g + bv_g^2 \quad . \quad . \quad . \quad (11\text{–}14)$$

than by (11–13). Hence if the input voltage is sinusoidal and is given by $V_m \sin pt$, then, ignoring the quiescent current, i_Q,

$$\begin{aligned} i_a &= aV_m \sin pt + bV_m^2 \sin^2 pt \\ &= bV_m^2/2 + aV_m \sin pt - bV_m^2/2 \,.\, \cos 2pt \end{aligned}$$

The first term in this result is the increase in the mean d.c. current resulting from the rectifying property of the triode. The last term

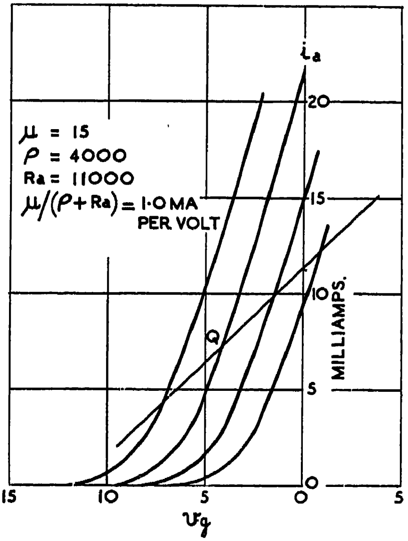


FIG. 11–10

shows that because of the non-linearity of the characteristic a second harmonic has been introduced. With a linear characteristic, we have $i_a = av_g$ and thus $a = \mu/(\rho + R_a)$. Hence

$$i_a = bV_m^2/2 + \frac{\mu V_m}{\rho + R_a} \sin pt - bV_m^2/2 \,.\, \cos 2pt \qquad (11\text{–}15)$$

where μ and ρ are evaluated at the origin Q, Fig. 11–10. It will be noted that the amplitude of the second harmonic is proportional to b. Differentiating (11–14) twice

$$\frac{d^2 i_a}{dv_g^2} = 2b$$

from which it is apparent that the amplitude of the second harmonic is proportional to the curvature of the dynamic characteristic.

In order to estimate the amount of harmonic distortion it is necessary to draw the dynamic characteristic. We may note here that if μ and ρ are constant the slope of this characteristic is given by $i_a/v_g = \mu/(\rho + R_a)$. Hence if the current through the valve is known in the absence of any applied voltage, the dynamic characteristic is simply obtained by drawing a straight line of slope $\mu/(\rho + R_a)$ through the quiescent or Q point as in Fig. 11–10. Considering the non-linear case, the Q point must first be determined, this being effected with the assistance of a family of curves

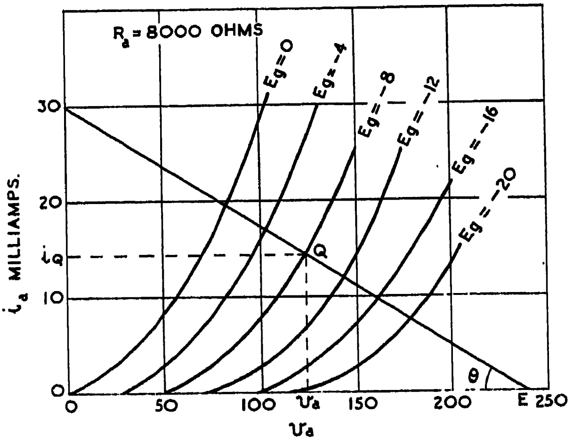


FIG. 11–11

known as anode characteristics. As will be seen on referring to Fig. 11–11, these characteristics give anode current against anode volts for various negative grid-cathode potentials. If E is the d.c. high-tension supply to the valve the anode voltage must be less than E because of the volt drop on R_a. Hence

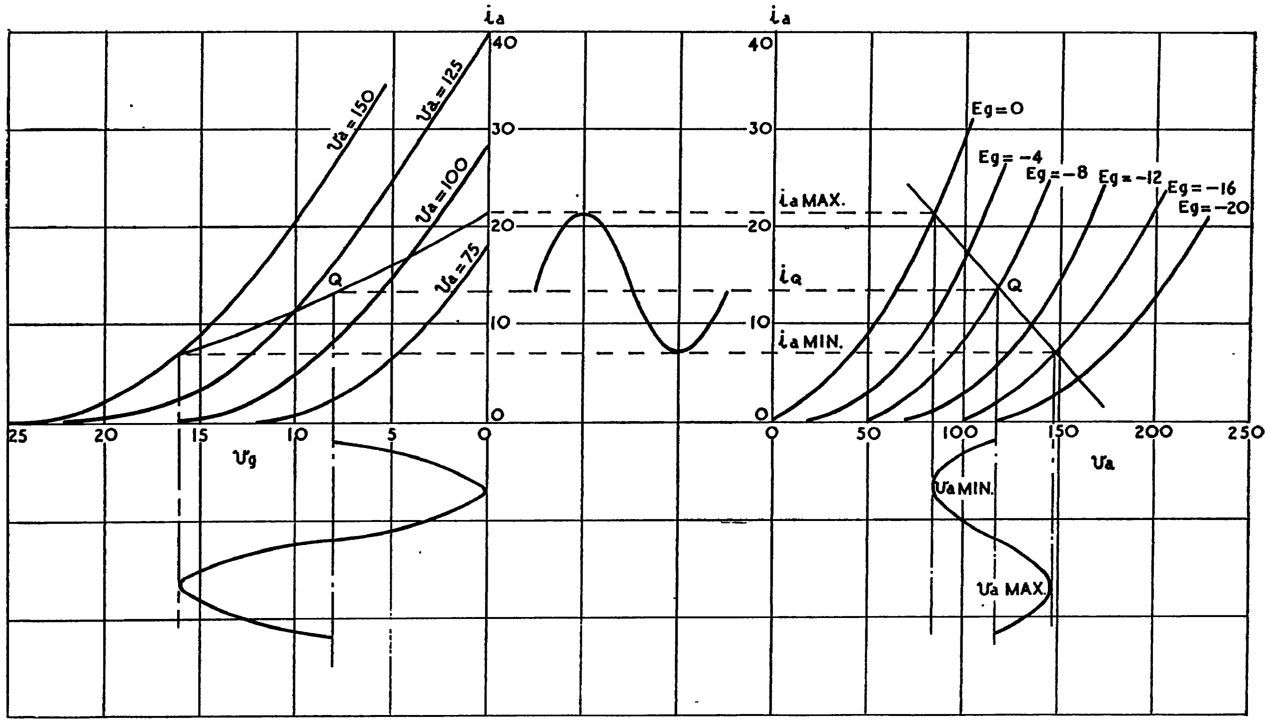
$$v_a = E - i_a R_a$$

$$R_a = \frac{E - v_a}{i_a}$$

and

$$i_a = \frac{E - v_a}{R_a}$$

Hence i_a is a linear function of v_a and is shown by the straight line drawn across the anode characteristic of Fig. 11–11. This line is known as the *load* line and its slope is given by $R_a = \cot \theta$. Evidently the Q point for any particular value of v_g is determined by



i_a
i_a
40
30
20
10
0
40
30
20
10
0
$v_a = 150$
$v_a = 125$
$v_a = 100$
$v_a = 75$
Q
Q
$E_g = 0$
$E_g = -4$
$E_g = -8$
$E_g = -12$
$E_g = -16$
$E_g = -20$
i_a MAX.
i_Q
i_a MIN.
25
20
15
10
5
0
v_g
0
50
100
150
200
250
v_a
v_a MIN.
v_a MAX.

FIG. 11–12

the intersection of the load line and the anode characteristic for this value of v_g.

Having determined the Q-point on the anode characteristics we may now determine the dynamic characteristic for a given input voltage. To effect this the anode and mutual characteristics are drawn side by side, as shown by Fig. 11–12. Corresponding values of i_a and v_g up to the limits of the variation of the latter are now transferred to the mutual characteristics in the manner shown. Then, drawing a curve through the derived co-ordinates on the latter, the dynamic characteristic results.

Being in possession of the dynamic characteristic and assuming it to be represented by (11–14), plus a quiescent term, it is noted from Fig. 11–12 and (11–15) that when $pt = \pi/2$

$$i_{a\,max} = i_Q + bV_m^2 + \mu V_m/(\rho + R_a)$$

also when $pt = -\pi/2$

$$i_{a\,min} = i_Q + bV_m^2 - \mu V_m/(\rho + R_a)^*$$

where i_Q is the value of the quiescent current. From these two equations

$$\frac{\mu V_m}{\rho + R_a} = \frac{i_{a\,max} - i_{a\,min}}{2}$$

$$\frac{bV_m^2}{2} = \frac{i_{a\,max} + i_{a\,min} - 2i_Q}{4}$$

which give the amplitudes of the fundamental and the second harmonic. The percentage second harmonic distortion is then

$$\frac{100bV_m^2/2}{\mu V_m/(\rho + R_a)}$$

Although the assumption that the dynamic characteristic may be expressed by an equation of second degree is valid for the majority of purposes, it is inadequate should the grid swing be such that the value is operated over the very curved portions of the characteristics, or should the grid go positive and upper bend curvature occur. In this case the dynamic characteristic is better represented by a cubic equation of the form

$$i_a = av_g + bv_g^2 + cv_g^3$$

Replacing v_g by $V_m \sin pt$, there results after simplification

$$i_a = bV_m^2/2 + (aV_m + 3cV_m^3/4) \sin pt$$
$$- bV_m^2/2 \,.\, \cos 2pt - cV_m^3/4 \sin 3pt$$

---

* $i_{a\,max}$ and $i_{a\,min}$ are here taken to represent total instantaneous values.

When $pt = \pi/2$

$$i_{a\,max} = i_Q + bV_m^2 + (aV_m + 3cV_m^3/4) + cV_m^3/4$$

Also $$i_{a\,min} = i_Q + bV_m^2 - (aV_m + 3cV_m^3/4) - cV_m^3/4$$

Adding $$\frac{bV_m^2}{2} = \frac{i_{a\,max} + i_{a\,min} - 2i_Q}{4}$$

which gives the increase in d.c. current and the amplitude of the second harmonic.

Subtracting

$$i_{a\,max} - i_{a\,min} = 2(aV_m + 3cV_m^3/4) + 2cV_m^3/4$$

and $$(aV_m + 3cV_m^3/4) = \frac{i_{a\,max} - i_{a\,min}}{2} - cV_m^3/4 \quad . \quad (11\text{–}16)$$

In order to obtain the amplitude of the third harmonic, i.e. $cV_m^3/4$, a tangent is now drawn to the dynamic characteristic at Q as shown by Fig. 11–13. Now, if the dynamic characteristic were a straight line, the limits of the anode current would be AC. Thus, in this case the amplitude of the fundamental would be $\mu V_m/(\rho + R_a)$ or AB. Substituting in (11–16)

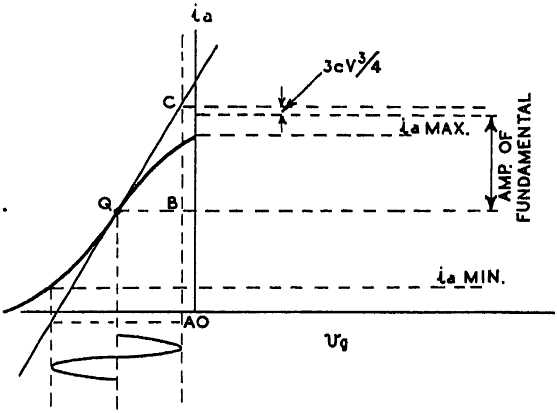


Fig. 11–13

$$(AB + 3cV_m^3/4) = \frac{i_{a\,max} - i_{a\,min}}{2} - cV_m^3/4$$

and $$\frac{cV_m^3}{4} = \frac{i_{a\,max} - i_{a\,min} - 2AB}{8}$$

which gives the amplitude of the third harmonic. Also, as the amplitude of the actual fundamental is $(AB + 3cV_m^3/4)$, we have for this

$$\frac{3(i_{a\,max} - i_{a\,min}) + 2AB}{8}$$

Generally speaking, the two foregoing analyses will cover most cases of harmonic distortion occurring in practice. However, if a more detailed study of the output waveform is desired, the wave may be constructed from the dynamic characteristic and then analysed by Fourier's method.

LOAD LINE WITH REACTIVE LOAD

In the event of the anode load being a pure resistance it has been shown that the locus of the anode current and voltage is a straight line, known as the load line. When the anode load is inductively reactive this is no longer true, the locus then being an ellipse. This is because the total anode voltage and current are, respectively, given by

$$v_Q - V_{a\,max} \sin pt \qquad . \quad . \quad . \quad (11\text{–}17)$$

and

$$i_Q + I_{a\,max} \sin (pt + \theta) \qquad . \quad . \quad (11\text{–}18)$$

where θ is the angle of phase displacement between V_a and I_a. If i_a is plotted against v_a for various values of pt an ellipse will result if $\theta \neq 0$. If $\theta = 0$ then the result is a straight line passing through the point i_Q, v_Q, the slope of the line being $v_Q/i_Q = R = \cot \theta$, as before.

In order to construct the ellipse for a given case, the load line is first drawn across the anode characteristics, as in Fig. 11–14, and the Q point determined. We then have

$$(v_Q + v_a) + (i_Q + i_a)R_a + L_a \frac{d(i_Q + i_a)}{dt} = E$$

or

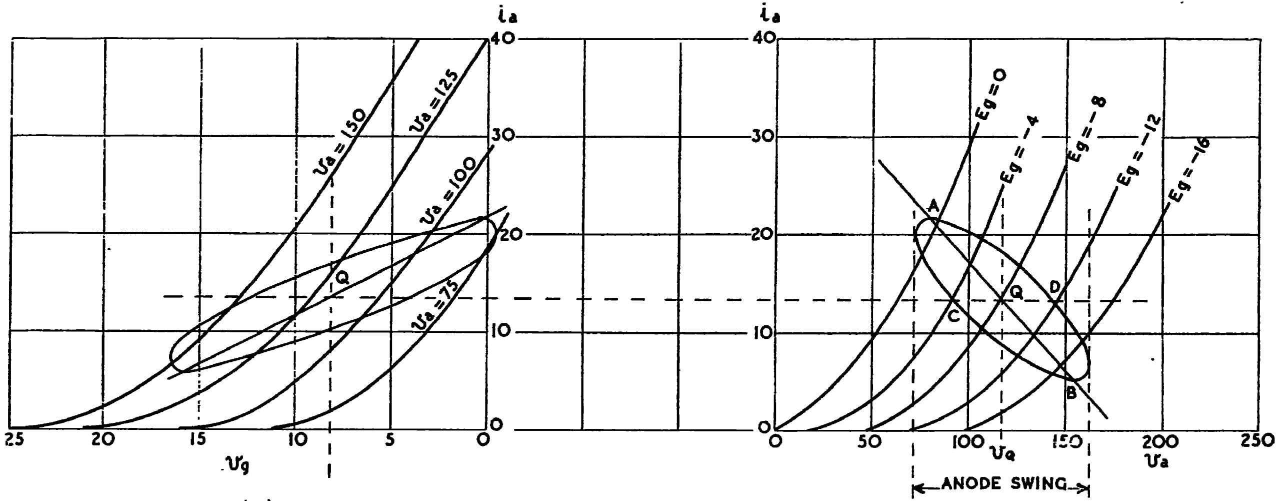
$$v_Q + v_a = E - (i_Q + i_a)R_a - L_a \frac{d(i_Q + i_a)}{dt}$$

Now. when $(i_Q + i_a)$ is a maximum or minimum,

$$v_Q + v_a = E - (i_Q + i_a)R_a$$

and hence under this circumstance the current and voltage co-ordinates lie on the load line at A and B. The maximum value of the current is, of course, given by

$$i_Q + \frac{\mu V_m}{\sqrt{(\rho + R_a)^2 + (pL_a)^2}}$$


i_a
40
30
20
10
0
25
20
15
10
5
0
v_g
$v_a = 150$
$v_a = 125$
$v_a = 100$
$v_a = 75$
Q
i_a
40
30
20
10
0
0
50
100
150
200
250
v_Q
v_a
$E_g = 0$
$E_g = -4$
$E_g = -8$
$E_g = -12$
$E_g = -16$
A
B
C
D
Q
ANODE SWING

(b)

(a)

FIG. 11-14

Again, when $i_a = 0$

$$v_Q + v_a = E - i_Q R_a - L_a \frac{di_a}{dt}$$

and $L_a di_a/dt$ is a maximum. In these circumstances the points C and D lie on a horizontal line passing through the Q point. Also

$$E - i_Q R_a = v_Q$$

so that

$$v_a = - L_a \frac{di_a}{dt}$$

and at C and D the anode voltage is $L_a di_a/dt$ volts smaller and greater than v_Q. The value of $L_a di_a/dt$ is, of course,

$$I_{a\,max} X_a = \frac{\mu V_m X_a}{\sqrt{(\rho + R_a)^2 + (pL_a)^2}}$$

Further points on the ellipse may be found from (11–17) and (11–18) by assuming arbitrary values for pt and thus determining various corresponding values of $(v_Q + v_a)$ and $(i_Q + i_a)$. θ is, of course, given by

$$\cos \theta = \frac{R_a}{\sqrt{R_a^2 + X_a^2}}$$

As the points on the load line of Fig. 11–12 were transferred to the mutual characteristics to form the dynamic characteristic, so also may the points on the ellipse of Fig. 11–14 be transferred. In this case, however, the dynamic characteristic is also an ellipse, as shown by Fig. 11–14 (*b*).

Power Amplification

The subject of amplification has so far been approached from the viewpoint of magnifying the input voltage only. However, the purpose of many amplifiers is to deliver power to some system such as an aerial, loudspeaker, or relay. Hence the output of an amplifier must now be considered from the aspect of *power* rather than that of voltage. Amplifiers are generally classified as Class A, AB, B, or C. Those previously discussed are of Class A, and this type as applied to power amplification will be dealt with first. It may be briefly stated that Class A amplifiers are those in which the anode and grid potentials are such that they operate over the appreciably linear portions of the characteristics, thus introducing a minimum of harmonic distortion. Of course, where resistance exists in the anode circuit of voltage amplifiers, power is developed therein, but in power amplifiers this resistance is usually of such value that the power is as large as possible consistent with absence of distortion. It follows that power-amplifier circuits are fundamentally similar to those employed for voltage amplification, and hence in analysing the former we may employ the methods already used for studying the characteristics of the latter.

Considering a triode with a resistance R_a in the anode circuit, the current is given by

$$I_a = \frac{\mu V_g}{\rho + R_a}$$

and thus the power supplied to R_a is

$$P = I_a^2 R_a = \left(\frac{\mu V_g}{\rho + R_a}\right)^2 R_a \qquad . \quad . \quad (11\text{–}19)$$

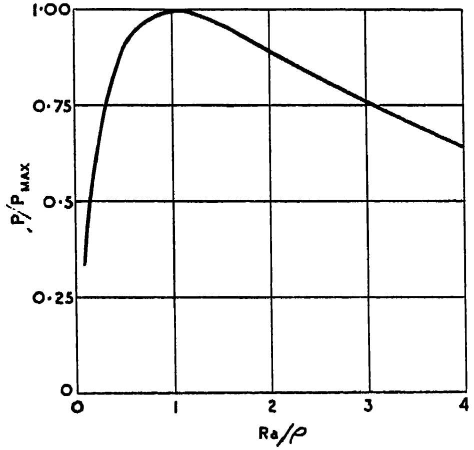


FIG. 11–15

If μ, ρ, and V_g are assumed to be constant, then maximum power will be developed in R_a when $dP/dR_a = 0$. Differentiating,

$$\frac{dP}{dR_a} = \left(\frac{\mu V_g}{\rho + R_a}\right)^2 - \frac{2R_a \mu^2 V_g^2}{(\rho + R_a)^3} = 0$$

or

$$\rho + R_a = 2R_a$$

and

$$R_a = \rho$$

Hence for maximum power, $R_a = \rho$ and under these conditions

$$P_{max} = \left(\frac{\mu V_g}{2\rho}\right)^2 \rho = \frac{\mu^2 V_g^2}{4\rho} \qquad . \quad . \quad (11\text{–}20)$$

In order to consider the effect on P of a departure from $R_a = \rho$, we may plot P/P_{max} against R_a/ρ. From (11–19) and (11–20)

$$\frac{P}{P_{max}} = \frac{4\rho R_a}{(\rho + R_a)^2} = \frac{R_a/\rho}{(1 + R_a/\rho)^2}$$

Reference to Fig. 11–15 shows that the loss of power from P_{max} is small, providing $0{\cdot}5\rho < R_a < 2\rho$.

From (11–19) it will be noted that in order that large powers may be developed μ and V_g should be large and ρ low. Hence, valves for power amplifiers usually have relatively low values of ρ and high values of μ. Also valve design is such that large grid swings may be handled without distortion occurring.

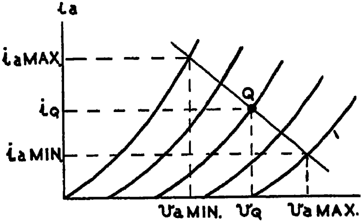


FIG. 11–16

GRAPHICAL DETERMINATION OF POWER

With the assistance of a set of anode characteristics the power may be determined graphically. Referring to Fig. 11–16 the load line is drawn and the Q-point determined. Now if the dynamic characteristic corresponding to the load line of Fig. 11–16 may be represented by an equation of second degree then, as shown on page 354, the amplitude of the fundamental component of the current is

$$\frac{i_{a\,max} - i_{a\,min}}{2} = I_{a\,max}$$

The fundamental power developed in the load is $I_{a\,max}V_{a\,max}/2$ and hence

$$P = \frac{(i_{a\,max} - i_{a\,min})\,(v_{a\,max} - v_{a\,min})}{8}$$

In previously determining the conditions for maximum power it was assumed that μ and ρ were constant. Although this may be approximately true for μ it is not for ρ, and hence, as we have already seen, the power will be distorted. The investigation of conditions to give maximum undistorted power is evidently a matter of some importance and must now be considered. The method of approach to this problem is to assume that the mutual characteristics of a triode are equidistant parallel straight lines, providing the anode current does not fall below some minimum value $i_{a\,min}$. In this case the dynamic characteristic will be a straight line with its lower extremity terminating on the line AB, as shown by Fig. 11–17. Below AB lower bend curvature commences. Now if the grid swing is such that grid current flows, this will impose a load on the output circuit of the previous valve, thus tending to introduce distortion into the input voltage, and hence the output voltage, of the amplifying valve. It follows that the grid voltage and steady bias should be such that the grid voltage never becomes positive. This results in the upper extremity of the dynamic characteristic terminating on the line Oi_a as shown. Static characteristics for anode voltages $v_{a\,max}$ and $v_{a\,min}$ are now drawn through A and C respectively. Midway between these characteristics is v_Q, the Q-point voltage, depending on R_a. The power output is proportional to

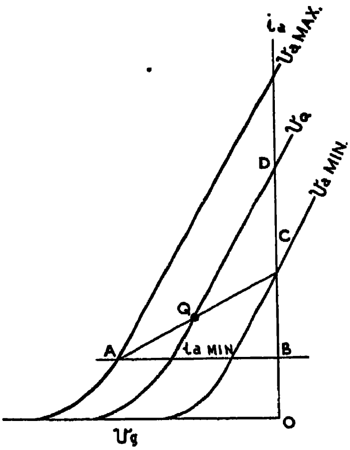


FIG. 11–17

$$(v_{a\,max} - v_{a\,min})\,(i_{a\,max} - i_{a\,min})$$

or to
$$(v_Q - v_{a\,min})CB$$

or to
$$CD\,.\,CB$$

As D and B are fixed, this product will be a maximum when $CD = CB$. Now

$$\begin{aligned} CD &= (v_Q - v_{a\,min})/\rho \\ &= (v_{a\,max} - v_{a\,min})/2\rho \end{aligned}$$

and thus the output will be a maximum when

$$(v_{a\,max} - v_{a\,min})/2\rho = CB = (i_{a\,max} - i_{a\,min})$$

or when
$$(v_{a\,max} - v_{a\,min})/(i_{a\,max} - i_{a\,min}) = 2\rho$$

But the left-hand side of this equation is equal to R_a and hence the maximum output occurs for

$$R_a = 2\rho$$

and
$$\begin{aligned} P_{max} &= \left(\frac{\mu V_g}{\rho + 2\rho}\right)^2 2\rho \\ &= \frac{2\mu^2 V_g^{\,2}}{9\rho} \end{aligned}$$

OPTIMUM VALUE OF GRID BIAS

It is necessary to determine the value of the grid bias in order that the maximum undistorted power may be obtained. We have

$$I_a = \mu V_g/(\rho + 2\rho)$$
$$= \frac{\mu V_g}{3\rho}$$

and $$V_g = \frac{3\rho I_a}{\mu}$$

Also $$I_a = (i_{a\,max} - i_{a\,min})/2\sqrt{2} = (v_{a\,max} - v_{a\,min})/4\rho\sqrt{2}$$

Hence $$E_g = \sqrt{2}\,V_g = \frac{3(v_{a\,max} - v_{a\,min})}{4\mu}$$

where E_g is the value of the steady d.c. bias required.

Parallel-feed

When the anode load resistance is placed directly in the anode circuit, the quiescent current passes directly through this resistance. This, of course, leads to a loss of power because the only power which is usually applied to some purpose is that produced by variations of the anode current about the quiescent point. Also it is sometimes desirable to prevent the direct component of the anode current from passing through the load. In view of these considerations the choke-capacitance arrangement of Fig. 11–18 is frequently employed in which the direct component passes through a choke. Because of the relatively high reactance of the choke over its working range it passes but a negligible portion of alternating current, the bulk of this passing through the load resistance R_a. The function of C is to block the d.c. voltage from R_a while passing practically unimpeded the alternating voltage. It follows that over the working range of frequencies the reactance of L must be high compared with the resistance of R_a, while that of C must be low. Departure from these conditions, say at low frequencies, will, of course, introduce frequency distortion.

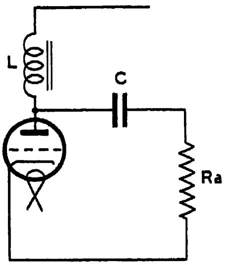


FIG. 11–18

The operating conditions when employing a parallel-feed system are shown by Fig. 11–19. In this case we have a static as well as a dynamic load line. The former is determined by the resistance of the choke and, as this resistance is normally low, the static load line will be almost vertical as shown. The Q-point is fixed by the intersection of this load line with the anode characteristic corresponding to the particular value of grid bias employed. v_Q will, of course, be approximately equal to E, the d.c. supply voltage. Over the working frequency range the alternating component of the anode current will flow in R_a and hence the dynamic load line passes through Q with a slope given by $R_a = \cot \theta$.

It is evident from the nature of the parallel-feed system that it cannot be employed for very low frequency work. As the frequency decreases, the impedance of L falls until it reaches such a value that the shunting effect on R_a causes the amplification to suffer a rapid decline. In general a parallel-feed amplifier is not suitable for distorted waveforms or transients because it tends to amplify the high-frequency components of the signal to a greater extent than the low-frequency ones. This, naturally, results in the output voltage wave not being an exact copy of the input wave.

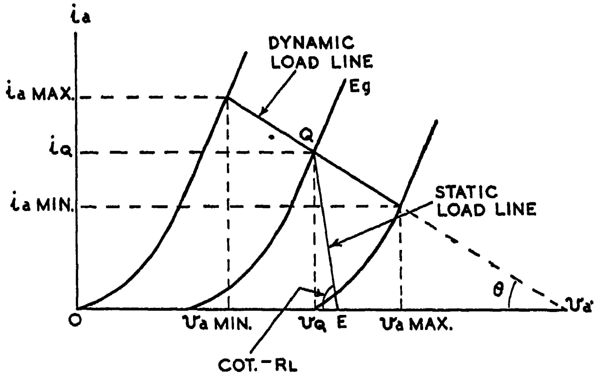


FIG. 11-19

Impedance Matching

In certain instances the load resistance is of fixed value and cannot be adjusted to give maximum undistorted power in the anode circuit. A typical case is, of course, the moving-coil loudspeaker, where the speech coil may have an effective resistance of no more than about 10 ohms. Evidently if this were connected directly in the anode line very little power would be developed. In practice adequate power is developed in the speech coil by matching its impedance to the slope resistance of the valve by means of a matching transformer, a typical circuit being shown by Fig. 11–20.

Now on page 344 it was shown that a resistance R_s in the secondary circuit of a transformer is equivalent to a resistance of value R_s/n^2 in the primary circuit. Hence if n is fractional, i.e. if a step-down transformer is employed, the effective value of R_a will be R_s/n^2. Hence by a suitable value of n, the transformer turns ratio, R_s/n^2 may be made equal to 2ρ or any other desired value. As with the parallel-feed circuit, a distinction must be recognized between the static and dynamic load lines for a transformer coupled load. In this case the static line is established by means of the transformer primary resistance, and the dynamic line by means of $R_a = R_s/n^2$. Parallel-feed may be employed with a transformer-coupled load as previously shown on page 347. Either an anode-feed resistance or a choke may be employed, and if the latter the circuit of Fig. 11–21 results.

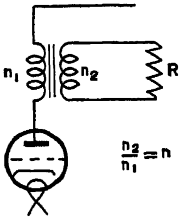


FIG. 11–20

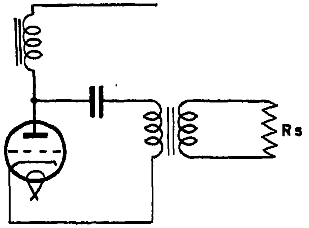


FIG. 11–21

Anode-circuit Efficiency

Under quiescent conditions the d.c. supply to a valve is largely dissipated in the form of heat at the anode, this heat being produced by electron bombardment. When the grid is excited by an alternating voltage, power is generated in the anode load, and this results in a reduction in the anode dissipation. The power developed in the load is, of course, supplied from the d.c. source, and we must now consider the relations existing between the alternating and direct current powers.

The power supplied by the d.c. source is Ei_Q, while that absorbed by the *static* load resistance, R, is $i_Q{}^2R$. If P_a is the average power dissipated at the anode and I_aV_a that in the load, then

$$Ei_Q = i_Q{}^2R + I_aV_a + P_a$$

and $$P_a = Ei_Q - i_Q{}^2R - I_aV_a$$

But $$E = v_Q + i_QR$$

and hence $$P_a = i_Qv_Q - I_aV_a$$

In the event of the load not being a pure resistance, the load power is $I_aV_a \cos \theta$, where θ is the angle of phase displacement between I_a and V_a. Thus, more generally, we have

$$P_a = i_Qv_Q - I_aV_a \cos \theta$$

From this result it will be at once appreciated that the greater the a.c. power developed in the load the smaller is the rate of energy dissipation at the anode. Obviously the latter is a maximum in the absence of grid excitation and decreases by exactly the same amount as the output power increases.

The efficiency of the anode circuit is defined as

$$\frac{\text{a.c. power output}}{\text{d.c. power input}} = \eta_a = \frac{I_aV_a \cos \theta}{Ei_Q}$$

If the overall efficiency of the valve is desired, then to the d.c. power input must be added the power dissipated in the grid and heater circuits. In this case the efficiency is given by

$$\frac{I_aV_a \cos \theta}{Ei_Q + P_h + P_g}$$

where P_h and P_g are, respectively, the rates of energy dissipation in the heater and grid circuits.

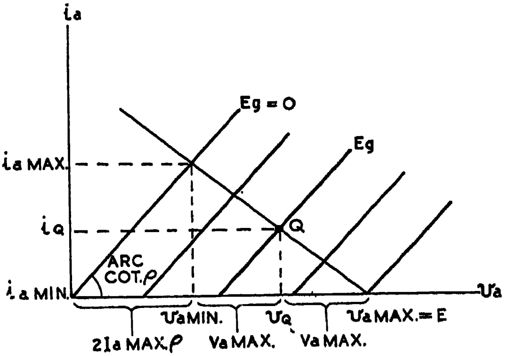


FIG. 11–22

In order to obtain some idea of the anode-circuit efficiencies possible with triodes it will be assumed that the anode characteristics consist of equidistant parallel straight lines, as shown by Fig. 11–22. This, of course, is not in accord with practice but, nevertheless, will permit limiting values of the efficiencies to be determined. The grid swing will be taken to be such that it causes $i_{a\,min}$ to be zero and $v_{g\,max} = 0$. In this case

$$i_Q = i_{a\,min}$$

and

$$\eta_a = \frac{I_aV_a}{Ei_Q}$$

$$= \frac{I_{a\,max}V_{a\,max}}{2EI_{a\,max}} = 0{\cdot}5\,\frac{V_{a\,max}}{E}$$

Now in the case of a series-fed load

$$E = 2V_{a\,max} + 2I_{a\,max}\,\rho \quad \text{(see Fig. 11–22)}$$

Hence

$$\eta_a = \frac{0{\cdot}5V_{a\,max}}{2(V_{a\,max} + I_{a\,max}\,\rho)}$$

$$= \frac{0{\cdot}25}{1 + \dfrac{I_{a\,max}\,\rho}{V_{a\,max}}}$$

$$= \frac{0{\cdot}25}{1 + \dfrac{\rho}{R_a}}$$

$$= \frac{25}{1 + \dfrac{\rho}{R_a}} \text{ per cent}$$

From this result it will be seen that the maximum possible efficiency with a series-fed load is 25 per cent, this necessitating R_a to equal infinity. In the case of maximum power output $R_a = \rho$ and $\eta_a = 12{\cdot}5$ per cent. For maximum undistorted power $R_a = 2\rho$ and $\eta_a = 16{\cdot}6$ per cent. In practice, figures somewhat less than these are obtained, as it is not possible to utilize the entire load line without curvature of the characteristics and distortion occurring. Considering the case of the shunt- or transformer-coupled load

$$E = V_{a\,max} + 2I_{a\,max}\,\rho$$

and

$$\eta_a = \frac{0{\cdot}5V_{a\,max}}{V_{a\,max} + 2I_{a\,max}\,\rho}$$

$$= \frac{0{\cdot}5}{1 + \dfrac{2\rho}{R_a}}$$

$$= \frac{50}{1 + \dfrac{2\rho}{R_a}} \text{ per cent}$$

Thus, the maximum possible efficiency with a shunt-fed load is twice that of a series-fed load. For maximum power $\eta_a = 16{\cdot}6$ per cent, while for maximum undistorted power $\eta_a = 25$ per cent.

Amplification with Pentodes

The power developed in the load resistance of a triode is $v_a i_a$ and the ratio $v_a i_a / v_g^2$ is termed the *power sensitivity* of the triode.*

---

* Because the power is proportional to v_g^2. See (11–19).

Providing the valve is operating on the linear portions of the characteristics, we may write

$$\frac{v_a i_a}{v_g^2} = \frac{dv_a di_a}{dv_g dv_g}$$

$$= \frac{dv_a}{dv_g} \cdot \frac{di_a}{dv_g} = \mu g_m$$

The mutual conductance of triodes and pentodes is of the same order of magnitude, whereas the amplification factor of pentodes is much larger than that of triodes. Thus, it follows that the power sensitivity of pentodes exceeds that of triodes, this rendering the pentode a more suitable valve for power work than a triode. In particular, a pentode will give the same power as a triode for a much smaller signal voltage.

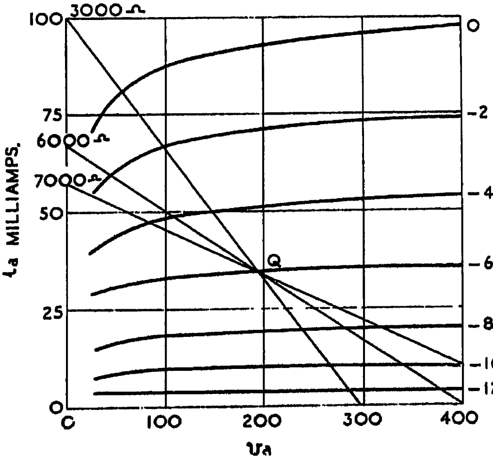


FIG. 11-23

As we have already seen, the anode characteristics of pentodes differ considerably from those of triodes, and this results in the power analyses previously made for triodes being invalid for pentodes. In particular, the dynamic characteristic of the pentode does not approximate to a parabola, nor is maximum undistorted output obtained when $R_a = 2\rho$. In the case of triodes, increasing R_a increases the linearity of the dynamic characteristic, thus leading to a decrease in distortion. With pentodes, however, increasing R_a beyond a certain value increases the curvature of the dynamic characteristic, thus increasing the distortion. A result of this is that, in practice, R_a is very much less than ρ; in fact, in some cases R_a may not exceed 10 per cent of ρ.

In order to determine the value of the anode resistance for minimum distortion, the Q-point and grid swing must be known. The Q-point is, of course, determined by the values of E and E_g. As the power that may be developed increases with E it is desirable to employ the maximum voltage permitted by the manufacturer's rating of the valve. Referring to Fig. 11–23, which gives the anode characteristics for a typical power pentode, the maximum anode voltage and anode dissipation of this valve are, respectively, 200 volts and 7 watts. Assuming E to be approximately 200 volts, then E_g may be taken as — 6 volts, which limits the anode dissipation under quiescent conditions to about 7 watts. Under these circumstances $i_Q = 35$ mA, and with a 6000-ohm load and a grid swing of 4 volts it will be noted that $I_{a\ max} = I_{a\ min} = 25$ mA. Also $V_{a\ max} = V_{a\ min} = 150$ volts. As the maximum and minimum values of the oscillatory components of the anode current and voltage are equal, this shows an absence of distortion. Now, let a load of line corresponding to $R_a = 3000$ ohms be drawn. In this case $I_{a\ max} = 32$ mA and $I_{a\ min} = 25$ mA, $V_{a\ max} = 110$ volts, $V_{a\ min} = 90$ volts. Hence distortion has been introduced. Again, let a load line corresponding to 7000 ohms be drawn. Then $I_{a\ max} = 22$ mA, $I_{a\ min} = 25$ mA, $V_{a\ max} = 210$ volts, $V_{a\ min} = 160$ volts. Thus, it appears that 6000 ohms constitutes the optimum load if distortion is to be avoided.

In order to compare the power in the three cases we have

$$P = (i_{a\ max} - i_{a\ min})\,(v_{a\ max} - v_{a\ min})/8$$

and substituting values there results

R_a ohms	P watts
3000	1·3
6000	1·88
7000	1·48

Thus, it appears that the highest power results from a load that renders the distortion a minimum. In the case of the 6000-ohm load the anode efficiency is

$$\eta_a = \frac{1{\cdot}88}{7{\cdot}0} = 27 \text{ per cent}$$

Also

$$\mu g_m = \frac{1{\cdot}88}{(4/\sqrt{2})^2} = 0{\cdot}235 \text{ watt per volt}$$

From the foregoing it appears that, when the Q-point and grid swing are known, the optimum load may be determined by successive approximations; i.e. a series of load lines is drawn through the Q-point and that which gives a minimum of distortion then determined. In effecting this it may be noted that because of the parallelism of the characteristics to the right of the Q-point there is practically no variation in $I_{a\ min}$, as R_a is varied. Hence $I_{a\ min}$ should first be found and a load line then selected which gives $I_{a\ max} = I_{a\ min}$.

Push-pull Amplification

By the employment of push-pull amplification several important advantages may be secured over the forms of amplification already discussed. This system employs pairs of valves arranged in the manner shown by Fig. 11–24. By means of a centre-tapped transformer two equal voltages differing in phase by 180° are applied to the grids of the valves. The output transformer is also centre-tapped, and as the quiescent currents of the two valves flow through the two halves of the winding in opposite directions, it follows that d.c. magnetization of the transformer core is absent. Thus, the primary inductance of the transformer is a maximum which at once shows one advantage of push-pull working.

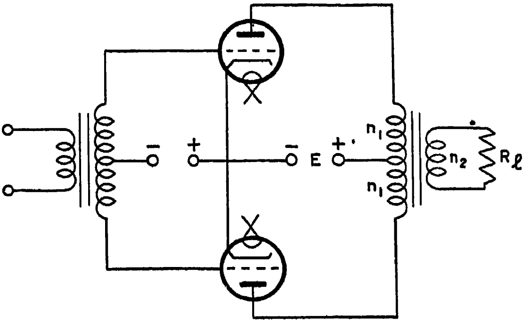


FIG. 11–24

Assume now that two voltages represented by $v_{g1} = V_m \sin pt$ and $v_{g2} = -V_m \sin pt$ are applied to the grids of the valves. Then on the assumption that the anode current/grid voltage relation is given by $i_a = av_g + bv_g^2$,

$$i_{a1} = aV_m \sin pt + bV_m^2 \sin^2 pt$$

$$i_{a2} = -aV_m \sin pt + bV_m^2 \sin^2 pt$$

or

$$i_{a1} = aV_m \sin pt + \tfrac{1}{2}bV_m^2 - \tfrac{1}{2}bV_m^2 \sin 2pt \quad . \quad (11\text{–}21)$$

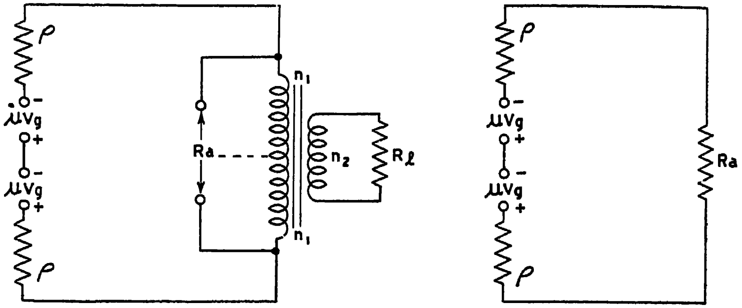
$$i_{a2} = -aV_m \sin pt + \tfrac{1}{2}bV_m^2 - \tfrac{1}{2}bV_m^2 \sin 2pt \quad . \quad (11\text{–}22)$$

Now when equal currents flow through the two halves of the output transformer in opposite directions, the resultant magnetic effect is zero. Thus, terms with the same signs in (11–21) and (11–22) produce no effect, and from this it follows that the second harmonics neutralize each other, as do also the increases in mean current, $\frac{1}{2}bV^2$. The terms with opposite signs directly assist each other so that the effective e.m.f. across the total primary winding of the output transformer is

$$(i_{a1} - i_{a2})R_a = \frac{\mu R_a}{\rho + R_a} 2V_m \sin pt$$

From this analysis it is apparent that a far more linear response may be obtained from a push-pull amplifier than from the amplifiers previously discussed.

A further advantage is as follows. If the source of d.c. supply to the valve anodes has appreciable internal resistance, a fluctuating load current will cause fluctuations in the d.c. voltage. These fluctuations will be impressed on the valves, but will be without effect on a push-pull system as the algebraic sum of the anode currents flowing to the valves is constant.



(a) FIG. 11–25 (b)

POWER OUTPUT OF PUSH-PULL AMPLIFIERS

Assuming a push-pull amplifier is operated under Class A conditions, the equivalent circuit may be represented by Fig. 11–25 (*a*), which may be further modified to (*b*). The mid-point connexion may be omitted because the fundamental component of the current does not pass through this. Referring to (*a*) the value of R_a is given by

$$R_a = \left(\frac{2n_1}{n_2}\right)^2 R_l \quad . \quad . \quad . \quad . \quad . \quad (11\text{–}23)$$

The anode current is

$$I_a = \frac{2\mu V_g}{2\rho + R_a}$$

or

$$I_a = \frac{\mu V_g}{\rho + R_a/2}$$

The power developed in R_a is

$$P = I_a^2 R_a = \left(\frac{\mu V_g}{\rho + R_a/2}\right)^2 R_a$$

$$= 2\left(\frac{\mu V_g}{\rho + R_a/2}\right)^2 \frac{R_a}{2}$$

$$= \left(\frac{\mu V_g}{\rho/2 + R_a/4}\right)^2 \frac{R_a}{4}$$

From this result it is apparent that a Class A push-pull amplifier may be regarded as consisting of a single valve with a slope resistance of $\rho/2$ working into an anode resistance equal to $R_a/4$. In order to find the conditions under which the power is a maximum, P must be differentiated with respect to R_a and the result equated to zero. Thus

$$\frac{dP}{dR_a} = \frac{\mu^2 V_g^2(\rho/2 + R_a/4)^2 - 2R_a\mu^2 V_g^2(\rho/2 + R_a/4)/4}{4(\rho/2 + R_a/4)^4} = 0$$

from which

$$\frac{R_a}{4} = \frac{\rho}{2}$$

or

$$R_a = 2\rho$$

COMPOSITE ANODE CHARACTERISTICS

The foregoing results are only correct providing μ and ρ are constant. Frequently, however, push-pull amplifiers are operated with grid swings of such magnitude that μ and ρ vary. In this case the operating conditions must be analysed by a graphical method. As the valves work in pairs it is desirable to devise a set of mutual characteristics for a pair as a single unit, such characteristics being known as *composite static characteristics*.

To effect this an ideal output transformer is first assumed. In this case, if v is the voltage across each half of the primary, then the load voltage is

$$v_l = \frac{n_2}{n_1} v \quad . \quad . \quad . \quad . \quad (11\text{–}24)$$

At every instant the primary ampere-turns are exactly balanced by the secondary ampere-turns and thus

$$n_1 i_{a1} - n_1 i_{a2} = n_2 i$$

where i_{a1} and i_{a2} are the *total* instantaneous anode currents. Also

$$v_l = iR_l = (i_{a1} - i_{a2}) \frac{n_1}{n_2} R_l \quad . \quad . \quad (11\text{–}25)$$

From (11–24) and (11–25)

$$v = (i_{a1} - i_{a2})\left(\frac{n_1}{n_2}\right)^2 R_l$$

or, from (11–23),

$$v = (i_{a1} - i_{a2})\frac{R_a}{4} \quad . \quad . \quad . \quad . \quad (11\text{–}26)$$

These results show that the load current is proportional to $(i_{a1} - i_{a2})$ and hence the purpose of the composite characteristics is to show the relationship between $(i_{a1} - i_{a2})$ and v_a with the grid voltage as a parameter.

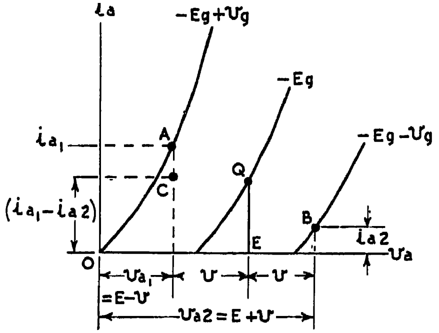


FIG. 11–26

Taking a set of anode characteristics for one valve of the pair, the Q-point of this valve is determined by the d.c. supply voltage E and the steady bias voltage $-E_g$, Fig. 11–26. Considering point A, this corresponds to $v_{a1} = (E - v)$ and $v_{g1} = (-E_g + v_g)$, the anode current being i_{a1}. Point B corresponds to $v_{a2} = (E + v)$ and $v_{g2} = (-E_g - v_g)$, the anode current being i_{a2}. Subtracting B from A gives a third point C (equal to $(i_{a1} - i_{a2})$) corresponding to an arbitrarily chosen value of v and a specified signal voltage v_g. It follows that the co-ordinates of C, i.e. $(i_{a1} - i_{a2})$, v_{a1} form a point on the composite characteristics. Assuming other values of v, and maintaining v_g constant, further points on the characteristic may be derived and a curve drawn through these points.

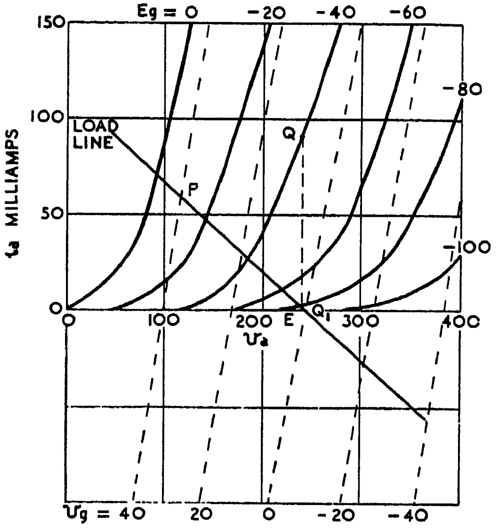


FIG. 11–27

By taking several different values of v_g a family of composite characteristics may be derived as shown by Fig. 11–27, which concerns a medium-size output triode. In this case the Q-point corresponds to $E = 240$ volts, $E_g = -40$ volts. It will be noted the characteristics form a series of parallel straight lines. The load line to be employed with composite characteristics is given by

$$v_{a1} = E - v$$
$$= E - (i_{a1} - i_{a2})\frac{R_a}{4}$$

which is the equation to a straight line passing through the point $v_{a1} = E$, $(i_{a1} - i_{a2}) = 0$, the slope being given by $R_a/4 = \cot\theta$. The intersection of the load line with the composite characteristics at P gives

$$(i_{a1} - i_{a2}) = i_p$$

and, from (11–25),

$$i = (i_{a1} - i_{a2})\frac{n_1}{n_2} = \frac{n_1}{n_2} i_p$$

Composite Dynamic Characteristics

In order to derive the composite dynamic characteristic of a push-pull amplifier, a similar process is employed to that on page 353, where the dynamic characteristic is derived from the anode characteristics. The fact that the composite mode characteristics are practically straight lines results, of course, in the composite

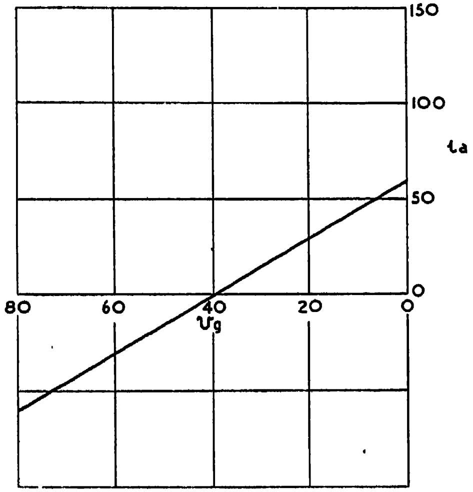


Fig. 11–28

dynamic characteristic being a straight line, as shown by Fig. 11–28, this being derived from Fig. 11–27. Hence it follows that a sinusoidal signal voltage results in an almost sinusoidal output voltage.

Class B Amplification

Class B push-pull amplification employs the same form of circuit as Class A, but conditions are so arranged that the quiescent current is very small or zero. These conditions may be brought about either by biasing ordinary valves to cut-off or by employing special valves having a very small quiescent current (or none) without the application of negative bias. The latter valves often consist of two triodes housed in a single envelope, the Mullard PM2BA and PM2B being examples. It follows that with Class B operation anode current mainly flows when the grids are driven positive, and hence grid current may be expected to be relatively large. That this is the case will be seen from Fig. 11–29, which concerns a class B valve. Because of the relatively high intake of the grids the previous valve is called upon to deliver appreciable power and must be of such character as to deliver this power without waveform distortion. For obvious reasons this valve is termed the *driver* valve.

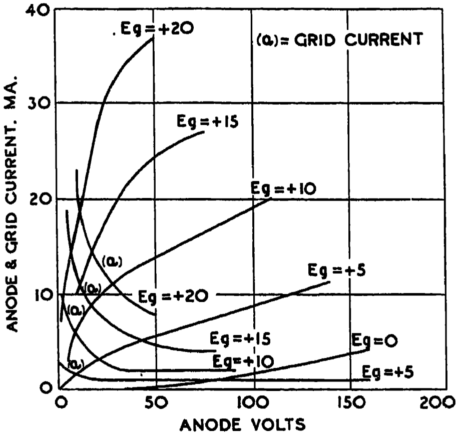


FIG. 11–29

The principal advantages of Class B operation are the small or non-existent demand on the d.c. supply when no signal is applied to the grid, and a high anode-circuit efficiency. The disadvantages are the necessity for a driver valve, greater distortion than Class A operation, and the inability to employ self-bias.

ANODE-CIRCUIT EFFICIENCY

In order to derive the anode-circuit efficiency the same assumption will be made as on page 365, i.e. that the anode characteristics are equidistant, parallel straight lines. It will be further assumed that the valves are biased to cut-off, that the grid swing does not extend beyond zero volts (see Fig. 11–30), and that the dynamic characteristic is a straight line. These assumptions are by no means justified in practice, but their adoption will, as before, lead to an upper limiting value for the anode-circuit efficiency.

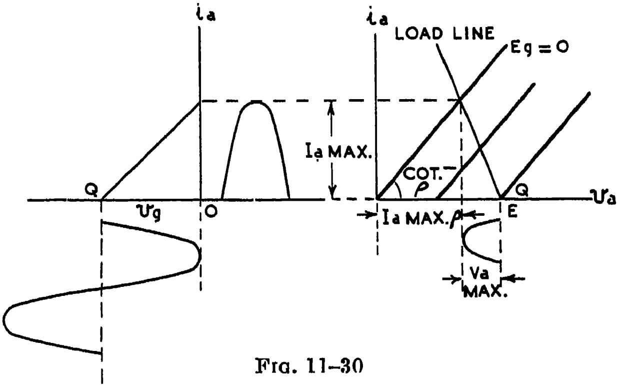


FIG. 11–30

Referring to Fig. 11–30, this shows the effect of a sinusoidal signal voltage on one of the two valves. It will be noted that the anode current consists of semi-sine waves, one of these occurring for half of each period followed by a completely quiescent half-period. In this respect each valve behaves in a similar manner to a half-wave rectifier. During the quiescent half-period of the first valve the anode current of the second valve describes a semi-sine wave exactly similar to that of the first valve. Hence these two successive half-waves produce the effect in the output transformer of a complete sinusoidal wave. The output power is given by

$$P = \left(\frac{I_{a\,max}}{\sqrt{2}}\right)^2 R_a$$

$$= \frac{I_{a\,max}^2 R_a}{2}$$

The mean anode current per valve is

$$\frac{1}{2\pi}\int_0^\pi I_{a\ max} \sin \theta d\theta$$

$$= \frac{I_{a\ max}}{\pi}$$

Hence the d.c. power is

$$2 \cdot \frac{I_{a\ max}E}{\pi}$$

the factor 2 occurring as there are two valves. The anode-circuit efficiency is

$$\eta_a = \frac{I_{a\ max}^2 R_a}{2} \bigg/ \frac{2I_{a\ max}E}{\pi}$$

$$= \frac{\pi}{4}\frac{I_{a\ max}R_a}{E} \quad . \quad . \quad . \quad . \quad (11\text{–}27)$$

Referring to Fig. 11–30,

$$E = I_{a\ max}\rho + V_{a\ max}$$

or, as $V_{a\ max} = I_{a\ max} R_a$

$$E = I_{a\ max}(\rho + R_a)$$

Substituting in (11–27)

$$\eta_a = \frac{\pi}{4}\frac{R_a}{\rho + R_a}$$

$$= \frac{0{\cdot}785}{1 + \rho/R_a}$$

$$= \frac{78{\cdot}5}{1 + \rho/R_a} \text{ per cent}$$

Comparing this result with that for Class A on page 366, it will be noted that Class B operative shows a considerable gain in anode-circuit efficiency. As previously shown, for maximum power output $R_a = 2\rho$, and hence in these circumstances $\eta_a = 52{\cdot}3$ per cent.

In order to obtain the composite characteristics of a Class B amplifier, exactly the same procedure is followed as when deriving those for Class A.* However, if the valves are biased to cut-off, the composite characteristics will coincide with the anode characteristics. If special valves requiring no grid bias are employed, there will be a small quiescent current. Hence in this case the composite characteristics will not coincide with the anode characteristics at low signal voltages, but the curves will tend to merge as the amplitude of the signal voltage increases. As the composite characteristics are not linear it follows that the dynamic characteristic is also non-linear. Of course, with Class B the curvature of the characteristic of one valve is no longer compensated by the curvature of the other, as with Class A working. Reference to Fig. 11–29 shows that the anode characteristics for a positive grid triode are similar to those for a pentode. Hence, in determining the value of the load resistance for a minimum of distortion, similar considerations as when dealing with a pentode load will apply.

THE DRIVER STAGE

As stated above, because of power consumption by the grid the previous value must supply power in a Class B system. From this it follows that the impedance of the driver valve circuit must not be too high or distortion of the input signal will result. Distortion is likely to arise because the relation between grid voltage and grid current is non-linear, as is evident from Fig. 11–29. Hence a non-sinusoidal voltage drop will be subtracted from the input voltage, which if the latter is sinusoidal, results in a non-sinusoidal voltage being applied to the grids of the Class B valves. To avoid this difficulty it is customary to employ a step-down transformer of low secondary resistance, in the manner shown by Fig. 11–31. For the valve of Fig. 11–29 the overall transformer ratio should be 1½ : 1 and the total resistance of the secondary should not exceed 400 ohms. Thus, the ratio for each secondary is 3 : 1 and the resistance not more than 200 ohms.

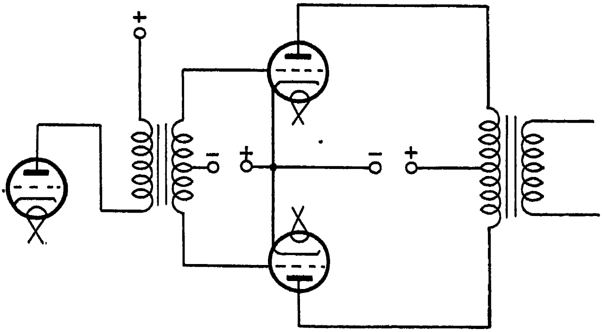


Fig. 11–31

Class AB Amplification

A form of amplification intermediate between Class A and Class B is Class AB. In this form the value of the negative bias is intermediate between the values employed for Classes A and B, with the result that the quiescent point of each value occurs at lower bend curvature. This allows a given output to be obtained with a smaller quiescent current than with Class A amplification, but with little more distortion because of the second harmonic eliminating properties of push-pull amplification. It follows that with Class AB amplification and a sufficiently large input voltage, lower bend cut-off may occur, thus causing the valve to conduct for less than a period but for more than half a period.

CHAPTER XII

FEED-BACK—VALVE OSCILLATORS

Negative Feed-back

By provision of suitable coupling between the output and input circuits of a valve it is possible to effect a considerable modification of its characteristics. The result of the coupling is to combine a portion of the output voltage with the input voltage, this process being termed *feed-back*. The degree of coupling may be either "tight" or "loose," according to whether the effect of feed-back is pronounced or slight. If the result of feed-back is to increase the original input voltage, it is said to be "direct," "positive," or "regenerative." If the input voltage is decreased, then the feed-back is said to be "inverse," "negative," or "degenerative." At first sight it might appear that all the advantages would be with positive feed-back, but, as will be shown later, considerable advantages are attached to negative feed-back. At present it may be briefly stated that the principal functions of positive feed-back are the amplifying of weak signals and the production of continuous oscillations. In the negative case feed-back may be employed for the reduction of frequency and harmonic distortion and improvement of amplifier stability.

Let a voltage $\mathbf{V}_g$ be applied to the input terminals of an amplifier and an output voltage $\mathbf{V}_a$ result. Following this, let a fraction α of the output voltage be fed back in such a manner that a voltage $\alpha\mathbf{V}_a$ is connected in series with $\mathbf{V}_g$ so as to reduce the input voltage. Then

$$\mathbf{V}_{gr} = \mathbf{V}_g + \alpha\mathbf{V}_a$$

where $\mathbf{V}_{gr}$ is the resultant of $\mathbf{V}_g$ and $\alpha\mathbf{V}_a$. Vector notation must be employed in this connexion because $\mathbf{V}_g$ and $\mathbf{V}_a$ are not necessarily in phase or antiphase. The gain of an amplifier is given by

$$\mathbf{M} = \frac{\mathbf{V}_a}{\mathbf{V}_{gr}}$$

hence

$$\mathbf{V}_{gr} = \frac{\mathbf{V}_a}{\mathbf{M}}$$

and

$$\mathbf{V}_a = \frac{\mathbf{M}\mathbf{V}_g}{1 - \alpha\mathbf{M}} \quad . \quad . \quad . \quad (12\text{–}1)$$

The resultant gain is given by

$$\mathbf{M}_r = \frac{\mathbf{V}_a}{\mathbf{V}_g}$$

$$= \frac{\mathbf{M}}{1 - \alpha\mathbf{M}} \quad . \quad . \quad . \quad (12\text{–}2)$$

Considering now the influence of feed-back on an amplifier, if $\alpha\mathbf{M}$ is large compared with unity (12–2) becomes

$$\mathbf{M}_r = -\frac{1}{\alpha}$$

and the gain depends only on the system producing feed-back. For example, if α is independent of frequency, the amplifier will be free from frequency distortion. If $\alpha\mathbf{M} \gg 1$, then $\mathbf{M}_r \ll \mathbf{M}$ and the gain is much smaller with negative feed-back than when this is not employed. Hence to secure the advantages attached to feed-back some loss of amplification must be sustained. However, this loss may be made good by an additional stage of amplification.

INFLUENCE ON HARMONIC DISTORTION

In addition to its beneficial influence on frequency distortion, feed-back may be employed to reduce non-linear, or harmonic, distortion. For example, suppose a second harmonic is produced by a valve in its output circuit. If $\mathbf{A}$ is the amplitude of this harmonic, this amplitude will appear in the output in the absence of feed-back. However, when feed-back is employed, the amplitude is, say, $\mathbf{A}_1$. Because of $\mathbf{A}_1$, the amplitude due to this and feed-back is $\alpha\mathbf{M}\mathbf{A}_1$, and hence the two components in the output are $\mathbf{A}$ and $\alpha\mathbf{M}\mathbf{A}_1$. It follows that the vector sum of these must equal $\mathbf{A}_1$ and hence

$$\mathbf{A}_1 = \mathbf{A} + \alpha\mathbf{M}\mathbf{A}_1$$

or

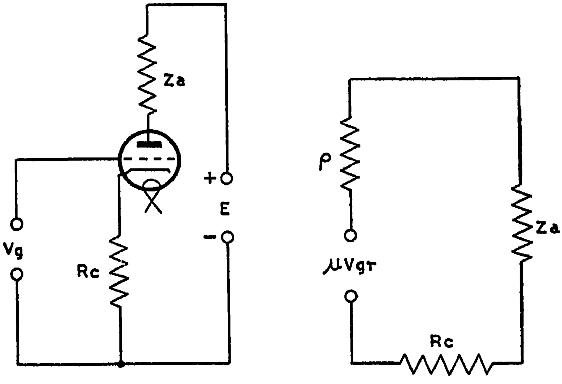
$$\mathbf{A}_1 = \frac{\mathbf{A}}{1 - \alpha\mathbf{M}}$$

This result at once shows that harmonic distortion may be reduced by the employment of negative feed-back. Thus, if $|1 - \alpha\mathbf{M}|$ is 10, the second harmonic distortion is only one-tenth of the value it would have in the absence of feed-back. As the magnitude of the harmonic depends on the amplifier output voltage the latter must be the same in both cases, i.e. with and without the feed-back. This means, of course, that V_g must be $|1 - \alpha\mathbf{M}|$ times as great in the presence of feed-back as in its absence.

In order to increase the input to the amplifier by the factor $|1 - \alpha\mathbf{M}|$ it is necessary to increase the amplification of the preceding stages or add a further amplifying stage. It is hardly necessary to state that these processes must not introduce additional distortion.

Negative Feed-back Circuits

Negative feed-back may be applied to an amplifier in either one, or a combination, of three basically different ways. Referring to Fig. 12–1 (*a*), this shows a "current" feed-back circuit, feed-back being applied by means of the cathode resistance R_c. In this case $\alpha = R_c/\mathbf{Z}_a$ and for large feed-back factors



(a) (b)

FIG. 12–1

$$\mathbf{M}_r = -1/\alpha = -\mathbf{Z}_a/R_c$$

Also $$\mathbf{V}_a = \mathbf{Z}_a\mathbf{V}_g/R_c$$

which is proportional to the load impedance. The load current is

$$\mathbf{I}_a = \frac{\mathbf{V}_a}{\mathbf{Z}_a} = \frac{\mathbf{V}_g}{R_c}$$

To determine I_a, we note from the equivalent circuit, Fig. 12–1 (*b*), that

$$I_a\mathbf{Z}_a + I_aR_c + I_a\rho = \mu\mathbf{V}_{gr}$$

But $$\mathbf{V}_{gr} = \mathbf{V}_g - I_aR_c$$

and hence $$I_a = \frac{\mu V_g}{\rho + (1 + \mu)R_c + \mathbf{Z}_a}$$

This result shows that the amplification factor of the valve is μ, but the slope resistance must be regarded as equal to

$$[\rho + (1 + \mu)R_c]$$

Thus, in this case feed-back increases the effective slope resistance.

A further method of producing feed-back is indicated by the circuit of Fig. 12–2, this being known as a "voltage" feed-back circuit. R is made large compared with $\mathbf{Z}_a$, while the reactance of C is small compared with R. The function of C is to isolate R from the d.c. supply voltage. Now from (11–1)

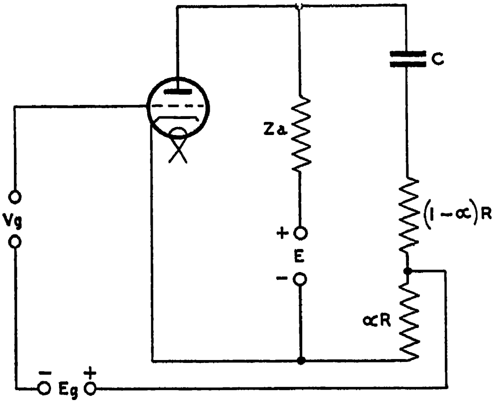


FIG. 12–2

$$\mathbf{M} = \frac{-\mu \mathbf{Z}_a}{\rho + \mathbf{Z}_a}$$

and as

$$\mathbf{M}_r = \frac{\mathbf{M}}{1 - \alpha \mathbf{M}}$$

$$\mathbf{M}_r = \frac{-\mu \mathbf{Z}_a}{\rho + (1 + \alpha\mu)\mathbf{Z}_a} \quad . \quad . \quad . \quad (12\text{–}3)$$

α in this case not being complex. Dividing the numerator and denominator of the last expression by $(1 + \alpha\mu)$, it may be written

$$\mathbf{M}_r = \frac{-\mu' \mathbf{Z}_a}{\rho' + \mathbf{Z}_a}$$

where $\mu' = \mu/(1 + \alpha\mu)$ and $\rho' = \rho/(1 + \alpha\mu)$

This expression indicates that feed-back of the type indicated by Fig. 12–2 gives conditions which would result from a valve having a slope resistance ρ' and an amplification factor μ' feeding into a load $\mathbf{Z}_a$, without feed-back being employed. It will be noted that both ρ and μ are reduced in the same proportion.

COMPOUND FEED-BACK

By simultaneously employing both current and voltage feed-back, what is known as compound feed-back is produced. The circuit for effecting this is shown by Fig. 12–3 and will be seen to be a combination of Figs. 12–1 and 12–2. As we have already found, the effect of current feed-back is to increase ρ to $\rho + (1+\mu)R_c$. Hence, combining this result with (12–3), we have

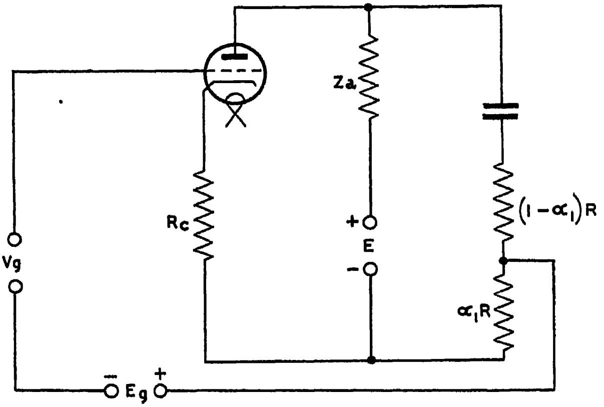


FIG. 12–3

$$\mathbf{M}_r = \frac{-\mu\mathbf{Z}_a}{\rho + (1+\mu)R_c + (1+\alpha_1\mu)\mathbf{Z}_a}$$

or

$$\mathbf{M}_r = \frac{-\mu'\mathbf{Z}_a}{\rho'' + \mathbf{Z}_a}$$

where $\mu' = \mu/(1+\alpha_1\mu)$ and $\rho'' = [\rho + (\mu + 1)R_c]/(1 + \alpha_1\mu)$.

From this result it will be noted that while μ is reduced, ρ'', i.e. the effective slope resistance, may be made larger than, equal to, or less than ρ by suitable choice of values for R_c and α_1.

Figs. 12–1, 12–2, and 12–3 refer only to basic feed-back circuits. In practice feed-back may be employed with more than one stage of amplification, the feed-back occurring from the last stage to the first or to any intermediate stage as desired. As an example of this, reference may be made to Fig. 12–4, which shows a two-valve resistance-capacitance coupled amplifier feeding back from the second valve to the first. Consideration of the potential changes throughout the circuit for a given polarity on the grid will show the feed-back to be negative.

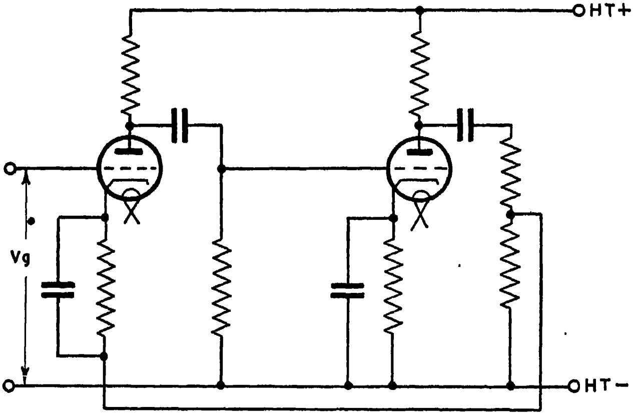


FIG. 12–4

Positive Feed-back

Under negative feed-back conditions $|1 - \alpha M| > 1$. However, if $|1 - \alpha M| < 1$ the feed-back is positive. An immediate result of this is that $M_r > M$ and thus positive feed-back may be employed for increasing the output of an amplifier for a given input voltage.

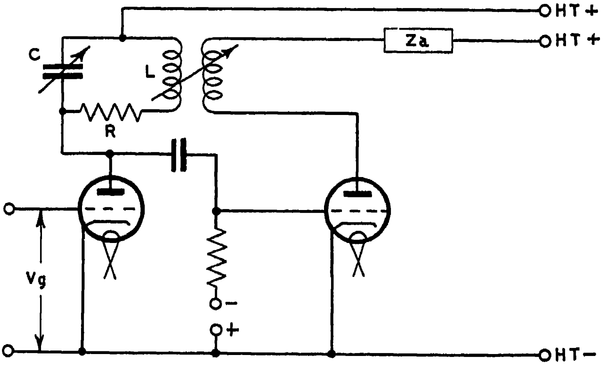


FIG. 12–5

Positive feed-back is frequently employed in connexion with the reception of wireless signals, where it is termed reaction. One method of application is shown by Fig. 12–5, where it will be noted that a coil in the anode circuit of the second valve is back-coupled to the tuned anode coil in the first valve. It has been stated previously that if the anode circuit is tuned to resonance it behaves as a pure resistance of value L/CR. Hence it is desirable that R shall be as small as possible if a high value of M is to be obtained. By coupling the anode coils in the manner indicated by Fig. 12–5, it may be shown that the effective value of R may be reduced below its normal value.

If I_1 is the current flowing round the tuned circuit, then the voltage applied to the grid of the second valve is

$$I_1/jpC$$

the resulting current in the anode circuit being

$$I_2 = \frac{\mu I_1}{jpC(\rho + \mathbf{Z}_a)}$$

If V is the e.m.f. acting in the tuned circuit, then

$$I_1[R + j(pL - 1/pC)] + jpmI_2 = V$$

or substituting for I_2

$$I_1\left\{\left[R + \frac{\mu m}{C(\rho + \mathbf{Z}_a)}\right] + j(pL - 1/pC)\right\} = V$$

If the anode circuit is tuned to resonance $pL = 1/pC$ and

$$I_1\left[R + \frac{\mu m}{C(\rho + \mathbf{Z}_a)}\right] = V$$

or

$$R + \frac{\mu m}{C(\rho + \mathbf{Z}_a)} = \frac{V}{I_1} \quad . \quad . \quad (12\text{–}4)$$

Now the left-hand side of this equation may be regarded as the effective resistance of the tuned anode circuit and, as m may be made either positive or negative, it is evident that R may be reduced and the amplification thus increased.

Valve Oscillators

In the foregoing example of the application of reaction, it has been shown that an increase in the voltage amplification factor results from a reduction in the effective resistance of the tuned anode circuit. It is evident from (12–4) that with m negative a possibility exists of making the effective resistance negative. In the event of this occurring, the amplifier passes into a state of self-oscillation. Without feed-back from the second valve an oscillation started in the LC circuit of the first valve quickly dies away because of the losses in R. However, if the energy can be supplied to L from L_1 at a rate equal to or greater than that at which it is dissipated in R, the oscillation will continue or even increase in amplitude.

To consider this matter in detail the conditions of the single valve circuit of Fig. 12–6 will be analysed. Let it be supposed that the amplification factor of this valve is sufficiently high that the value of the anode current is almost solely determined by the value of the grid potential. Then, according to (10–6),

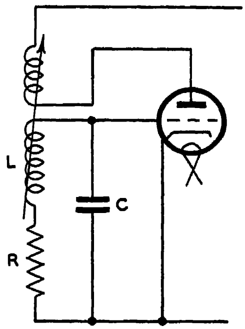


Fig. 12–6

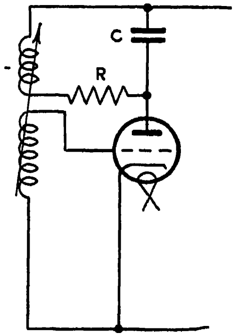


Fig. 12–7

$$i_a = g_m v_g$$

Assuming an oscillation to be started in the LC circuit, the conditions may be written

$$L\frac{di}{dt} + m\frac{di_a}{dt} + Ri + v_g = 0$$

where i is the oscillatory current. Now $i = C dv_g/dt$ and $i_a = g_m v_g$. Substituting, we may write

$$LC\frac{d^2v_g}{dt^2} + mg_m\frac{dv_g}{dt} + RC\frac{dv_g}{dt} + v_g = 0$$

or

$$\frac{d^2v_g}{dt^2} + \frac{1}{L}\left(R + \frac{mg_m}{C}\right)\frac{dv_g}{dt} + \frac{v_g}{LC} = 0$$

From this result we again see that because of coupling between the anode and grid circuits the effective resistance of the latter has been altered from R to $(R + mg_m/C)$. Again if the coil connexions are arranged so that m is negative, it is possible by varying m to make the effective resistance equal to zero. In this case an oscillation once started will persist at a frequency approximately equal to $1/2\pi\sqrt{LC}$. If the effective resistance becomes negative, then the oscillations will increase in amplitude.

In order to cause oscillations to occur some electrical disturbance is necessary, such as the closing of the switch in the anode circuit. Once started, the oscillations build up spontaneously to an amplitude the magnitude of which will be considered below. The fact that oscillations have started is conveniently indicated by an ammeter in the anode circuit. Because of the rectifying properties of the valve, there is usually an increase in the value of the mean anode current when oscillatory action commences.

The condition for oscillations to be maintained in the circuit of Fig. 12–6 may be regarded from another viewpoint. The e.m.f. induced by the anode coil in the grid coil is $mg_m dv_g/dt$, this being in phase with the current i. Hence the anode circuit introduces energy into the grid circuit at a rate $img_m dv_g/dt$, while R dissipates energy at a rate i^2R. Now $i = Cdv_g/dt$ and, for oscillations to be maintained, it is necessary that energy shall be supplied as rapidly as it is dissipated. Hence

$$Cmg_m\left(\frac{dv_g}{dt}\right)^2 \geq RC^2\left(\frac{dv_g}{dt}\right)^2$$

and

$$\frac{mg_m}{C} \geq R \quad . \quad . \quad . \quad . \quad (12\text{–}5)$$

or

$$R - \frac{mg_m}{C} \leq 0$$

as before.

TUNED ANODE OSCILLATOR

As the alternative to the tuned grid circuit of Fig. 12–6, the tuned anode circuit of Fig. 12–7 may be employed for the maintenance of oscillations. If I is the oscillatory current in L, then the losses are I^2R. For oscillations to be maintained energy must be supplied to the tuned anode circuit at a rate equal to I^2R. Now, when oscillating, the anode circuit behaves as a pure resistance of value L/CR and hence the anode current is

$$I_a = \frac{\mu V_g}{\rho + \dfrac{L}{CR}}$$

or

$$\frac{\mu pmI}{\rho + \dfrac{L}{CR}}$$

The voltage across C is pLI and the power supplied to the anode circuit is

$$I_aV_a = \frac{\mu pmI}{\rho + \dfrac{L}{CR}} \cdot pLI = \frac{\mu p^2 mLI^2}{\rho + \dfrac{L}{CR}}$$

For oscillations to be maintained we must have

$$\frac{\mu p^2 mLI^2}{\rho + \dfrac{L}{CR}} \geq I^2R$$

or, putting $p^2 = 1/LC$,

$$\frac{\mu m}{C\left(\rho + \dfrac{L}{CR}\right)} \geq R$$

and $$m \geq \frac{CR\rho + L}{\mu} \geq \frac{CR}{g_m} + \frac{L}{\mu} \quad . \quad . \quad (12\text{–}6)$$

POWER IN OSCILLATORY CIRCUIT

By coupling the inductance in the circuits of Figs. 12–6 and 12–7 to some other circuit, it is, of course, possible to draw power from the oscillator. At any instant the current supplied by the d.c. supply is $i_Q + I_{a\,max} \sin pt$, while the voltage between anode and cathode is $E + V_{a\,max} \sin (pt + \pi)$. Hence the instantaneous power dissipated as heat at the anode is

$$\begin{aligned} &(i_Q + I_{a\,max} \sin pt)[E + V_{a\,max} \sin (pt + \pi)] \\ &= (i_Q + I_{a\,max} \sin pt)\,(E - V_{a\,max} \sin pt) \\ &= Ei_Q + EI_{a\,max} \sin pt - i_Q V_{a\,max} \sin pt - I_{a\,max}\, V_{a\,max} \sin^2 pt \end{aligned}$$

and the mean value of this over a cycle is

$$Ei_Q - \frac{I_{a\,max}\, V_{a\,max}}{2}$$

This is less than the power supplied by the d.c. source by an amount I_aV_a which is, consequently, the power available for maintaining oscillations.

MAXIMUM POWER

The maximum oscillatory power occurs when I_aV_a is a maximum. If i_s is the saturation current of the valve then $I_{a\,max} = i_s/2$; also $i_Q = i_s/2$. The maximum value of v_a is equal to E. These values are, of course, purely theoretical limiting values but will, nevertheless, give a limiting value to the maximum power and efficiency of the valve. It follows that

$$\frac{I_{a\,max} V_{a\,max}}{2} = \frac{Ei_s/2}{2} = \frac{Ei_s}{4}$$

The d.c. input power is

$$Ei_Q = Ei_s/2$$

and the efficiency is thus

$$\frac{Ei_s/4}{Ei_s/2} = 50 \text{ per cent}$$

ANODE TAPPING POINT

In order that a valve oscillator of the tuned anode type may deliver theoretical maximum power, a certain value of the ratio L/C must be achieved. We have,

$$V_{a\,max} = I_{a\,max} \times L/CR$$

or

$$E = \frac{i_s}{2} \cdot \frac{L}{CR}$$

and

$$\frac{L}{C} = \frac{2RE}{i_s}$$

Now for a given frequency the product LC is fixed and hence for maximum power the ratio L/C must be adjusted without altering the LC value. This is effected by placing part of L in series with C in the manner shown by Fig. 12-8. With this arrangement the

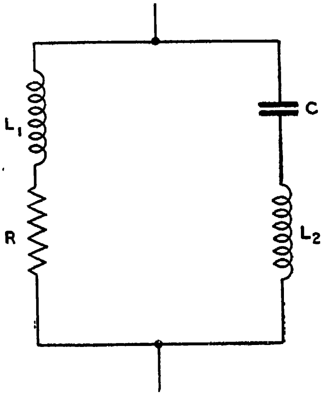


FIG. 12-8

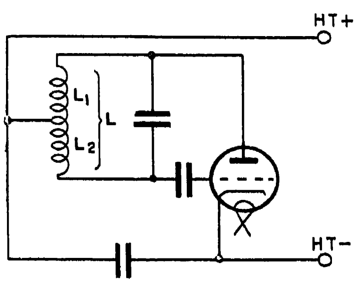


FIG. 12-9

frequency of oscillation is still given by $1/2\pi\sqrt{LC}$ but the combination of L_2 and C in series acts as a capacitance of increased value because the reactance of the combination is less than that of C acting alone. If the value of the effective capacitance is C_1, then

$$\frac{1}{pC_1} = \frac{1}{pC} - pL_2 = \frac{1 - p^2CL_2}{pC}$$

Now
$$p^2 = \frac{1}{L_1C_1} = \frac{1 - p^2CL_2}{L_1C}$$

or
$$p^2(L_1C + CL_2) = 1$$

and
$$p^2 = \frac{1}{(L_1 + L_2)C} = \frac{1}{LC}$$

which shows that the frequency is unaffected by the distribution of the inductance.

Again
$$C_1 = \frac{1}{p^2L_1} = \frac{LC}{L_1}$$

and
$$\frac{L_1}{C_1} = \frac{L_1^2}{LC}$$

From this result we see that the ratio of the inductance of the circuit to the effective capacitance is L_1^2/LC and hence the effective circuit resistance is

$$\frac{L_1}{C_1R} = \frac{L_1^2}{LCR}$$

Thus the effective circuit inductance and resistance have been decreased and the effective capacitance increased. It is now necessary to determine L_1. We have

$$V_{a\,max} = E = \frac{i_s}{2} \times \frac{L_1^2}{LCR}$$

from which
$$L_1 = \sqrt{\frac{2LCRE}{i_s}}$$

Again, if conditions are arranged for maximum efficiency, i.e. 50 per cent, the internal and external resistances of the oscillator must be equal. Hence

$$\frac{L_1^2}{LCR} = \rho$$

and
$$L_1 = \sqrt{R\rho LC}$$
$$= \sqrt{R\rho}/p$$

By placing part of the inductance in the capacitance arm the value of m to maintain oscillations is altered from the value given by (12–6). We now have

$$I_a = \frac{\mu pmI}{\rho + \frac{L_1^2}{LCR}}$$

and

$$V_a = pL_1I$$

Hence

$$\frac{\mu p^2 m L_1 I^2}{\rho + \frac{L_1^2}{LCR}} \geq I^2R$$

and

$$m \geq \frac{LCR\rho + L_1^2}{\mu L_1} \leq \frac{LCR}{L_1 g_m} + \frac{L_1}{\mu} \quad . \quad . \quad (12\text{–}7)$$

Of course, if all the inductance is in one branch of the circuit, $L_1 = L$ and (12–7) reduces to (12–6).

In addition to the two oscillator circuits previously described, a very large number of other valve oscillator circuits exist, so numerous in fact that space does not permit their description here. However, the circuits of Figs. 12–6 and 12–7 may be regarded as basic, the great majority of other circuits conforming to these in principle. This means that coupling exists between grid and anode circuits so that energy may be fed from one circuit to the other in order that oscillatory action may be maintained. Oscillator circuits need not consist of one valve only, but energy may be fed from the anode circuit of one valve to the grid circuit of another. In many cases the efficiency of a single-valve oscillator is of no great moment as it may be coupled to an amplifier, and it is the efficiency of the latter which is then of principal importance.

Grid Excitation

In the circuits of Figs. 12–6 and 12–7, the grid is excited by what is termed *mutual inductive coupling*. An alternative method of grid excitation is direct inductive coupling and is exemplified by the circuit of Fig. 12–9, sometimes referred to as the Hartley circuit. In this it will be seen that the inductance is divided into two parts with the junction connected via a blocking condenser to the cathode. There may or may not be mutual inductance between the two parts; actually this is immaterial. It is evident from the figure that the anode-grid and grid cathode voltages are 180° out of phase. The circuit may be re-drawn in the manner shown by Fig. 12–10, in which case it is evident that the circuit resembles that of Fig. 12–8 with L_2 taking the place of m. Hence the analysis applied to the circuit of Fig. 12–8 is applicable to that of Fig. 12–10 if L_2 takes the place of m. It follows that for the maintenance of oscillations

$$L_2 \geq \frac{LCR}{L_1 g_m} + \frac{L_1}{\mu}$$

An important point which may be noted with this circuit is that the anode-grid inter-electrode capacitance is in parallel with the tuned circuit capacitance, and thus does not act as an uncontrolled coupling capacitance.

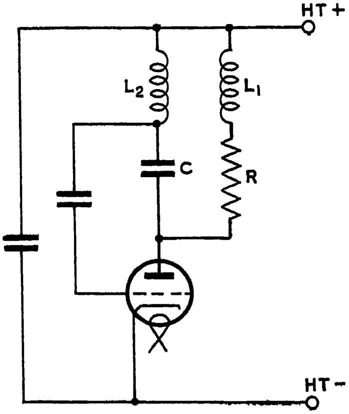


FIG. 12–10

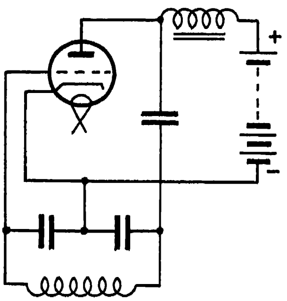


FIG. 12–11

A circuit employing direct capacitance coupling is shown by Fig. 12–11. This is sometimes referred to as the Colpitts circuit, and the capacity is so divided as to provide a voltage between grid and cathode sufficient to maintain oscillations.

FEED

As with amplifiers, oscillator circuits may employ either series or shunt feed. The Colpitts circuit shown has shunt feed, while the three previous circuits are series fed. The alternative shunt-feed circuits for the tuned-anode and Hartley circuits are shown by Figs. 12–12 and 12–13, respectively.

USE OF GRID BIAS WITH OSCILLATORY CIRCUITS

It has been shown that the maximum efficiency of a tuned-anode oscillator is 50 per cent. By negatively biasing the grid the quiescent current may be reduced and also the mean anode dissipation. This, of course, makes possible an increase in efficiency beyond 50 per cent. However, a detrimental effect of grid bias is the introduction of harmonics into the output. If the grid is so biased that when oscillations build up the lower curved portion of the i_a/v_g characteristic is involved, then evidently distortion will occur. A further point in connexion with the employment of grid bias is that the value of m must be greater to start and maintain oscillations. When bias is employed g_m is smaller than when it is not, and reference to (12–6) and (12–7) shows that, in consequence, m must be increased.

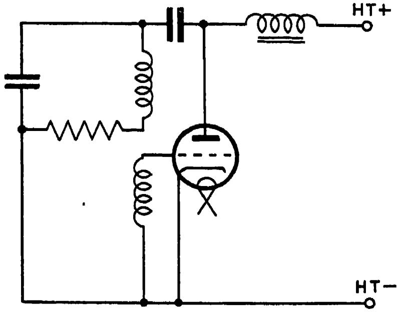


FIG. 12–12

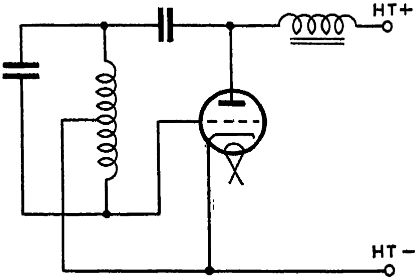


FIG. 12–13

Two methods of applying grid bias to an oscillator may be employed: (1) the grid may be self-biased as shown by the circuit of Fig. 12–14, or (2) a steady bias may be employed derived from an external source. Regarding the latter, if the bias is too great, g_m may become too small for oscillations to start. This objection does not, however, arise with method (1). With this method, when the valve is not oscillating, there is no bias applied to the grid. Hence g_m has its maximum value and, when an appropriate disturbance occurs, oscillations commence. During the positive half-cycles of oscillatory grid voltage the mean grid potential goes negative in the manner described in Chapter X. Hence the grid is only negatively biased while oscillations are occurring, and in this manner the biasing arrangements in no way affect the initiation of the oscillatory state.

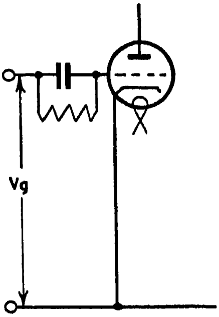


FIG. 12–14

AMPLITUDE OF OSCILLATIONS

So far the conditions necessary for starting and maintaining oscillations have been studied without closely considering the amplitude reached by the oscillations. In determining the efficiency of an oscillator, it was assumed that the oscillatory component of the anode current varied between 0 and i_s. However, in practice, this is not necessarily so. In the various cases considered it has been assumed in (12–5), (12–6), and (12–7) that g_m is constant. However, this is only true for the middle portion of the i_a/v_g characteristic. As upper and lower bend curvature are approached, Fig. 12–15, the value of g_m becomes less than at the quiescent point. Hence a point will be reached on the characteristic at which the inequalities (12–5), (12–6), and (12–7) become equalities. At this point the amplitude of the oscillations will no longer increase but will remain constant. Thus it will be appreciated that it is the non-linearity of the valve characteristics which is responsible for the stability of the oscillator. Regarded in another way, we may say with respect to the circuit of Fig. 12–6 that ultimately the value of g_m becomes such that the effective resistance of the circuit changes from negative to zero. Beyond this point the resistance becomes positive and hence the oscillation amplitude is limited by the value of g_m which causes the effective circuit resistance to change from negative to positive. With regard to the tuned-anode circuit, the losses in this circuit increase as the oscillation amplitude increases. Ultimately g_m decreases to such a value that the rate of introduction of energy into the anode circuit is just equal to the rate of dissipation of enegy therein. At this point the oscillation amplitude will cease to increase.

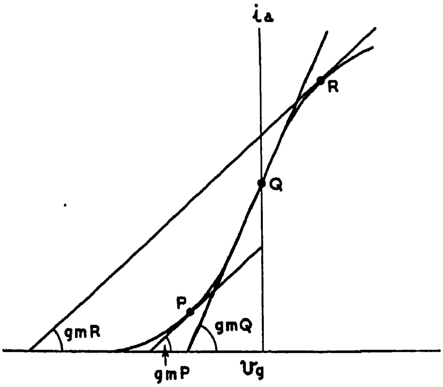


FIG. 12–15

It is evident that when a self-biased oscillatory circuit is employed this will also tend to stabilize the oscillator and limit the amplitude of oscillation. As described on page 393, with this system of bias, before oscillations commence, g_m is a maximum. On initiation of the oscillatory state the mean grid potential travels negative. Thus the working point moves down the characteristic resulting in a decrease in g_m.

Further Types of Oscillators

Before proceeding to describe several other oscillators it is desirable to offer some remarks concerning oscillators in general. Those so far described are of the feed-back type and actually consist of an amplifier feeding back part of the output to the input so that the amplifier provides its own excitation. Such oscillators may be also termed *sinusoidal oscillators* because, in general, they produce sinusoidal voltage and current waves. Where the output waves are markedly non-sinusoidal, the oscillator is frequently termed a *relaxation oscillator*, as the oscillations are characterized by an abrupt change, or relaxation, from one unstable state to another. The time-base circuits described in Chapter XVII generally comprise oscillators of this type.

A further form of oscillator is the *negative-resistance oscillator*. In such oscillators a circuit element is employed possessing a negative volt-ampere characteristic over some operating range. An example of this is the dynatron described in Chapter X. A valve operated in this manner possesses a negative slope resistance over a portion of its characteristic and this may counterbalance the positive resistance of an LC circuit in the anode lead, thus making continuous oscillations possible.

The basic circuit of a dynatron oscillator is shown by Fig. 12–16 and the valve characteristic by Fig. 12–17. Under oscillatory conditions the oscillatory circuit behaves as if possessing a resistance equal to L/CR. Hence if R_s is the value of the negative slope resistance of the valve, oscillations will commence providing $R_s \geq L/CR$. The slope of the load line is L/CR and under oscillatory conditions this must be less than R_s. Thus, referring to Fig. 12–17, oscillations are not possible with the load line (*a*), but are with (*b*). Although the load line will not, in general, coincide with the line R_s, nevertheless the current-voltage locus must be along the valve characteristic. Hence two current-voltage loci are necessary to describe the operation when $R_s > L/CR$. That of (*b*) gives the value of the *fundamental* component of the total current, while the valve characteristic gives the total current. The difference between these two currents at any point corresponding to a given anode voltage is the value of the harmonic current. Because of the low reactance of the condenser to harmonics of the fundamental, the harmonic components of the current mainly pass through the condenser. Thus, the alternating component of the anode voltage is appreciably sinusoidal, although the anode current may contain appreciable harmonics. The harmonic current for one value of the anode voltage is shown in Fig. 12–17 by P_1P_2.

It is possible for the oscillation amplitude to become such that the maximum and minimum of the valve characteristic are exceeded. In this case the slope resistance of the valve beyond these points is positive. Under these circumstances the average value of the slope resistance becomes less negative and the oscillation is stable at an amplitude at which the average value of the slope resistance is equal to L/CR.

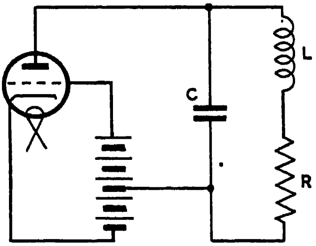


FIG. 12–16

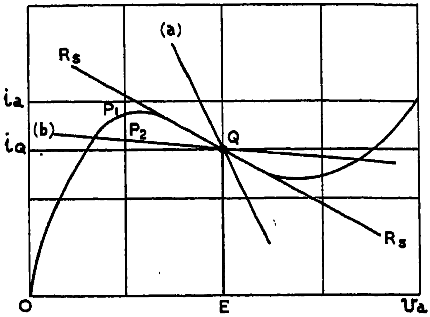


FIG. 12–17

An alternative method of treating the dynatron oscillator is to assume that the characteristic of Fig. 12–17 is that of a cubic equation, which may be expressed as

$$i_a = i_Q - \alpha v + \beta v^3 \quad . \quad . \quad . \quad (12\text{–}8)$$

The conditions in the LC circuit may be written

$$L\frac{d(i + i_a)}{dt} + R(i + i_a) + v = 0$$

$$= L\frac{di}{dt} + Ri + L\frac{di_a}{dt} + v = 0 \quad . \quad . \quad (12\text{–}9)$$

as i_a is small compared with i. Differentiating (12–8),

$$\frac{di_a}{dt} = \frac{di_a}{dv}\cdot\frac{dv}{dt}$$

$$= (-\alpha + 3\beta v^2)\frac{dv}{dt}$$

Substituting in (12–9) and putting $di/dt = Cd^2v/dt^2$,

$$LC\frac{d^2v}{dt^2} + (CR - L\alpha + L3\beta v^2)\frac{dv}{dt} + v = 0$$

or
$$\frac{d^2v}{dt^2} + \left(\frac{R}{L} - \frac{\alpha}{C} + \frac{3\beta v^2}{C}\right)\frac{dv}{dt} + \frac{v}{LC} = 0$$

An approximate solution of this equation is

$$v = 2\sqrt{\frac{\alpha - CR/L}{3\beta}}\cos pt$$

where $p = 1/\sqrt{LC}$. This result shows that oscillations will occur providing $\alpha \geq CR/L$, i.e. the reciprocal of the slope resistance at Q must be equal to or larger than CR/L. This, of course, is the same thing, as $R_s \geq L/CR$, the result previously found.

THE MAGNETRON OSCILLATOR*

Under suitable circumstances the cylindrically symmetrical system described on page 132 may be employed to generate ultra-high-frequency oscillations. In these circumstances the oscillator is termed a magnetron. In order to obtain the high frequencies of which the magnetron is capable, the external resonant circuit sometimes consists of two Lecher† wires arranged as shown by Fig. 12–18. The purpose of the three chokes is to prevent high-frequency currents from passing through the power supply circuits.

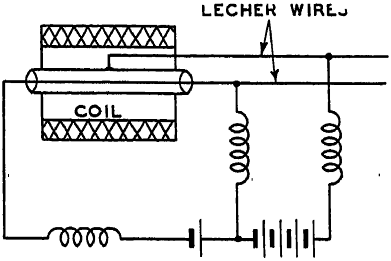


FIG. 12–18

Some uncertainty appears to exist concerning the principle by which oscillations are produced by the magnetron. However, it is possible that it functions in the following manner. On page 134 it is shown that the minimum value the anode voltage may have for current to flow is

$$V = \frac{H^2 q}{8m} r_c^2$$

---

* For recent magnetron developments see *An Introduction to Multi-Resonator Magnetrons*, R. Latham, A. H. King, L. Rushforth, *The Engineer*, 5th April, 1946, p. 310; *The Engineer*, 12th April, p. 331.

† *Electricity and Magnetism*, p. 459, S. G. Starling (Longmans, Green).

If, initially, the anode voltage is below this value, an increase above V will cause current to flow in the external circuit. This will cause the anode voltage to fall, whereupon, because of the weakened electric field, electrons emerging from the cathode will fail to reach the anode. Thus the current will cease and the anode voltage again rise. This re-initiates the current flow and the cyclic behaviour will be indefinitely repeated. The transit time of the electrons from cathode to anode is one-half the period of the cyclic behaviour.

BIBLIOGRAPHY

"Negative Feed-back," *Wireless World*, Feb., 1946, p. 41.

"Negative Feed-back and Hum," *Wireless World*, May, 1946, p. 142.

"Methods of Applying Negative Feed-back," *Electronic Engineering*, Nov., 1945, p. 770.

Theory and Design of Valve Oscillators, H. A. Thomas (Chapman and Hall).

"Cavity Magnetrons," *Wireless World*, May, 1946, p. 146.

CHAPTER XIII

ELECTRONIC RECTIFIERS

BEFORE the development of electronics, the derivation of a d.c. supply from an alternating source was invariably met by what was virtually some mechanical switching device. Among such devices may be mentioned the vibrating reed rectifier, motor-generator sets, synchronous converters, rotary converters, transverters, etc. Such rectifiers function by virtue of the motion of certain parts (such as a commutator) and thus this form of rectification is sometimes referred to as "dynamic." With the growth of electronics rectification by mechanical switching has largely given way to electronic methods, the various forms of the latter being termed static rectification.

Considerable advantages result from the employment of static rectifiers. Unlike rotary types, it is easily practicable to construct static rectifiers for almost any lower limit of capacity, a particularly important application in this respect being the operation of moving-coil instruments from a.c. sources. The particular advantages of electronic rectifiers may be stated as absence of moving parts, low maintenance costs, high and practically constant efficiency, negligible depreciation, and high momentary overload capacity.

Although many different forms of electronic rectifiers exist, the majority of applications employ either one of two basically different types. These are the dry-contact and valve types of rectifier. Amongst the former are the selenium and copper-oxide rectifiers, while the latter consists of the vacuum tube, the hot-cathode gas-filled, and mercury-arc rectifiers. The dry-contact, vacuum-tube, and gas-filled types are mainly used for relatively low power supplies, the mercury arc being employed for heavy power work. With respect to current and voltage limitations, single units of the foregoing group are capable of supplying from 0·1 mA for instrument operation up to 10,000 amp. for traction purposes, these ranges, of course, being capable of extension by the employment of units in parallel. With regard to voltage, single units of the vacuum type may be obtained to operate at almost any upper limit, while hot-cathode gas-filled and mercury-arc types have been constructed for 50 kV and 30 kV respectively. In the case of dry-contact rectifiers the voltage range is extended by placing a number of units in series. Evidently the range is practically unlimited. Besides a natural lower voltage limit for the foregoing rectifiers, an economical lower voltage limit exists. This is about 0·5 volt for copper-oxide, 1 volt for selenium, 10 volts for gas-filled hot cathode, and 100 volts for mercury arcs.

Asymmetrical Electrical Conductivity

The majority of conductors possess what may be termed symmetrical conductivity, i.e. the magnitude of the current flowing in either direction through the substance is independent of the polarity of the applied voltage. However, in certain conductors or arrangements of conductors, the magnitude of the current passing depends on the polarity of the applied voltage. In such cases the conductor, or arrangement, is said to possess asymmetrical conductivity, which means that it passes current more freely in one direction than the other. The direction of higher conductivity is termed the "permeable," "forward," or "conducting" direction, while the lower is termed the "impermeable," "reverse," or "non-conducting" direction. An example of these properties with which we have already dealt is, of course, the thermionic vacuum tube.

The essential property of all electronic rectifiers is that they possess asymmetrical conductivity. If an alternating e.m.f., whose mean value is zero, is applied to an asymmetrical system, it is evident that for one half-period more current will flow than for the other. Thus the mean value of the current will be different from zero and rectification will occur. Let it be assumed that the resistances in the forward and reverse directions are constant and respectively equal to r_f and r_r. If a load resistance R is placed in series with an asymmetrical conductor and a voltage given by $e \sin \theta$ applied, then the mean current is

$$\frac{1}{2\pi}\left[\int_0^{\pi} \frac{e \sin \theta d\theta}{r_f + R} + \int_{\pi}^{2\pi} \frac{e \sin \theta d\theta}{r_r + R}\right]$$

$$= \frac{e}{\pi}\left(\frac{1}{r_f + R} - \frac{1}{r_r + R}\right) \quad . \quad . \quad (13\text{–}1)$$

which will be different from zero, providing $r_f \neq r_r$. The voltage appearing on R is given by

$$\frac{eR}{\pi}\left(\frac{1}{r_f + R} - \frac{1}{r_r + R}\right) . \quad . \quad . \quad (13\text{–}2)$$

Commenting on (13–1) and (13–2) it is obviously desirable that r_r should be as large as possible and r_f as small as possible, the ideal condition being $r_r = \infty$, $r_f = 0$. With this condition, (13–1) and (13–2) become $e/\pi R$ and e/π respectively. In practice, of course, conditions are never ideal, but in many cases r_r is so high that it may be regarded as infinite. The ratio r_r/r_f is sometimes referred to as the rectification ratio.

Contact Rectifiers

Although numerous forms of contact rectifiers exist, only four may be said to be well known; these are the crystal, the electrolytic cell, the copper-oxide, and the selenium rectifiers. The first was, of course, extensively used in the early stages of radio-communication, while the last two are the only types which have been developed and may be said to be indispensable to the electrical industry. The electrolytic rectifier was at one time developed to the point that it was suitable for low-power work, but fell into disuse with the advent of metal rectifiers, i.e. the copper-oxide and selenium types.

THE ELECTROLYTIC RECTIFIER

Before passing to the detailed consideration of the two last-mentioned rectifiers, the electrolytic rectifier will be briefly considered as its principle is similar to that of all rectifiers of the contact type. If two electrodes of aluminium and iron are immersed in a solution of sodium bicarbonate ($NaHCO_3$) and a current passed so that the aluminium is the anode, the current will rapidly fall to a negligible figure. If the direction of the current is reversed, it will flow quite freely and electrolysis will take place in the usual manner. An explanation of these phenomena is the following. If aluminium is made an anode in a suitable solution, on switching on the current there will be a migration of OH ions towards the anode where they will combine with the aluminium to form a film of hydroxide on the surface. Successive ions on arriving find difficulty in penetrating this film and form a layer of gas in it, the film being porous. The function of the hydroxide film is thus largely mechanical; it provides a support for the gas layer. It will thus be seen that the presence of this gas layer gives the cell its high resistance when the aluminium is the anode. Rectification may, therefore, be attributed to the comparative ease with which the aluminium can send its free electrons through the gas layer when made cathode, the gas layer forming a somewhat effective barrier to the comparatively massive anions (negatively charged molecules) when the aluminium is made cathode.

In the light of the foregoing theory we may say that the electrolytic rectifier consists of an electrolyte which is poor in free electrons, an electrode rich in free electrons, and an insulating film which is formed on the electrode (anode) immediately after switching on the current. The metal used for the anode must be one which is easily oxidizable but reducible with difficulty. Thus, the action is electronic, electrons flowing from the aluminium through the film to the electrolyte, but not vice versa.

A reason for the limited use of electrolytic rectifiers is the fact that the anode film is easily reducible. During the reverse half-cycle the film tends to be destroyed, resulting in a relatively large reverse current and low reverse resistance. In some cases the author has found the rectification ratio to be no higher than 35 : 1.

The theory put forward for the action of the electrolytic rectifier is applicable to dry-contact rectifiers. In general, these consist of a combination of a metal carrying an insulating film and a poor conductor consisting of a binary metallic compound or mixture. The film is formed at the surface of the metal either by chemical or electrolytic action. Examples of such rectifiers are lightly oxidized aluminium and fused cupric sulphide; oxidized tantalum and lead peroxide; oxidized zinc and lead peroxide; copper and aluminium in contact immersed in ammonium sulphide, etc. All the foregoing are similar in that they consist of a good conductor, an insulating film, and a poor conductor.

In dealing with now-employed metal rectifiers it is customary to refer to the insulating film as the barrier or blocking layer. Assuming a potential difference is produced between the two conductors, the barrier layer will create a steep potential gradient between them. Assuming the polarity of the poor conductor is positive, the intense electric field at the surfaces of the conductors will cause auto-electronic emission from the better one in the manner described on page 184. The electrons therefore will be accelerated and acquire sufficient velocity to pass through the barrier layer and reach the poorer or semi-conductor. On reversing the potential a similar effect will occur, but its magnitude will be much smaller, due to the comparative lack of free electrons in the poorer conductor. In order to produce the most pronounced rectifying effect it is evident that the differences in the work functions, and free electron densities of the two conductors should be as great as possible, i.e. the good conductor should possess a low work function and the poor conductor a high one.

THE COPPER-OXIDE RECTIFIER

The copper-oxide rectifier dates from Grondahl's announcement in 1926 that a piece of copper having one side covered with a layer of cuprous oxide (Cu_2O), possesses marked asymmetrical conductivity. In order to produce rectifying units either discs or strips of copper are employed. The copper must possess a high degree of purity and it is essential that finish and cleanliness are perfect. However, in spite of this it appears that no relationship between chemical purity and asymmetry has yet been found. The copper must possess about 0·03 per cent of oxygen, as oxygen-free copper fails to form an oxide with the necessary adherence. It is found that non-metallic impurities such as sulphur, the halogens, and gases are of primary importance, while metallic impurities are of secondary importance only. The physical properties of the copper, such as hardness and grain size are of no importance and the most suitable way of testing copper for rectifier purposes is the construction of actual rectifiers from it.

THE OXIDATION PROCESS

Having obtained copper of the right degree of purity, the next process of great importance is its oxidation. As gaseous impurities are of primary importance, it is evident that oxidation must be carried out in a pure atmosphere. The copper blanks are heated in air at between 900° C. and 1020° C. for about ten minutes, during which time an oxide film approximately 0·1 mm. thick is formed. The copper is next transferred to a furnace at 600° C. and allowed to cool to this temperature. It is then removed and cooled to room temperature by quenching in oil or water. The purpose of both cooling processes is the production of a layer of cuprous oxide of the required resistance. At this stage the cuprous oxide takes the form of a hard crystalline layer, firmly adhering to the copper, but covered with a thin film of cupric oxide formed during the annealing. This film must be removed, and this is effected either by abrasion, sodium cyanide, or, more latterly, by mineral acid processes.

To describe the oxidation process in greater detail, it may be said that on first heating the copper from cold up to 400° C. cuprous oxide alone is formed. Above this temperature cupric oxide also commences to form, but tends to be confined to the surface due to the low porosity of the film. Due to this, the quantity of oxygen reaching the cuprous oxide is insufficient to oxidize it completely. As the temperature is increased, oxidation proceeds more rapidly, the surface film of cupric oxide continuing to exist until approximately

975° C., when it is reduced to cuprous oxide. Up to a temperature of 1020° C. the higher the temperature the lower the forward resistance of the rectifier. Above this temperature and below 900° C. the forward resistance tends to be excessive.

With regard to the annealing process, the second stage of this, i.e. cooling from 600° C., is the most important. The quicker this is accomplished the lower the rectifier resistance and hence, in practice, water quenching is most frequently employed.

THE BARRIER LAYER

The specific resistance of pure cuprous oxide is of the order of 10^8 ohms per cm.^{3}, which is a far higher figure than that found for the oxide present in rectifiers. The latter is of the order of 10^3 ohms per cm.^{3} or less. This is because cuprous oxide formed by direct oxidation of copper contains oxygen in excess of that demanded by the chemical formula. Actually it is found that the conductivity of the oxide depends on the pressure of the oxygen in the surrounding atmosphere. From these facts it may be considered that the copper-oxide rectifier consists of a good conductor, copper, a poor or semi-conductor, i.e. cuprous oxide having an excess of oxygen, and a barrier layer consisting of pure cuprous oxide. While oxidation is occurring, oxygen must be passing through the oxide already formed to reach the copper in order that oxidation may continue. When oxidation is discontinued, due to a reduction in temperature, the oxide will contain an excess of oxygen which must be retained so that the oxide may be partially conducting. The pressure of oxygen is greatest at the outer surface of the oxide, and it is probable that the oxygen immediately adjacent to the copper continues to combine with this, forming a thin layer of pure cuprous oxide. The latter will then form the barrier layer.

RECTIFIER RESISTANCE

The total resistance of any copper-oxide rectifying unit must comprise the resistances of the following—

1. The copper base.
2. The contact between copper base and oxide.
3. The oxide.
4. The contact between oxide and counter-electrode.
5. The counter-electrode.

With reference to (5), this electrode is a soft foil washer pressed on to the oxide in order to obtain a low-resistance contact. Further details of this electrode will be given later. It is evident that the resistances of (1), (4), and (5) may be neglected and we are thus left with those of (2) and (3). Referring to the latter, it is found that the cuprous oxide obeys Ohm's Law, but the specific resistance decreases from the outer surface towards the copper base. In practice the average specific resistance is of the order of a few hundred ohms per cubic centimetre. The oxide resistance, of course, varies with its thickness, but not linearly, due to the variation in specific resistance referred to above. Variation in the thickness of the oxide layer, therefore, affects the resistance of the rectifier, but principally in the forward direction, the effect on the high reverse resistance being negligible. The oxide resistance does not comprise the entire forward resistance, as the barrier layer forms the greater part of this.

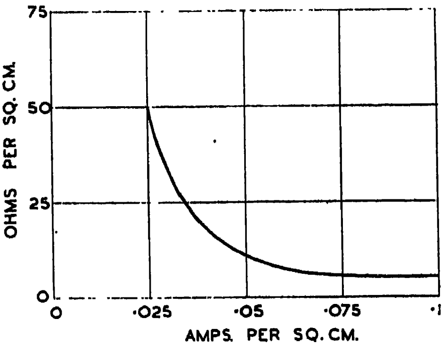


FIG. 13–1

Current and Temperature Effects

Copper-oxide rectifiers are influenced by current and temperature, the resistance falling with an increase in these quantities. Of the two, the current has the most pronounced effect, as shown by Fig. 13–1, which gives a forward resistance/current curve for a typical unit. As the oxide layer obeys Ohm's Law it is apparent that the effect of an increase in current is to reduce the resistance of the barrier layer. This is also true for the reverse direction as Fig. 13–2 shows. The influence of temperature on both forward and reverse resistances is indicated by the volt/ampere characteristics of Fig. 13–2. It will be noted that the resistance temperature coefficient is negative in both cases. Actually the temperature coefficient of cuprous oxide is found to be negative, but, in addition to this, it is evident that the barrier-layer coefficient is of the same type.

Electrical Properties of Copper-oxide Rectifiers

As indicated by Fig. 13–2, it is evident that both forward and reverse resistances are dependent on the applied voltage and that in consequence the rectifier does not follow Ohm's Law. In addition to the variation due to voltage, the rectifier resistance also varies with time. This latter effect can be divided into reversible changes known as "creep," and irreversible changes known as "ageing." Referring to Fig. 13–2, it will be noted that the forward conductivity of a rectifying unit is very low until the voltage reaches a value of about 0·25 volt. At this figure a sharp increase in current occurs with an increasing tendency in the characteristic to linearity. It will be seen that the reverse current rapidly increases with increasing reverse voltage, but does not exhibit the sharp bend in curvature of the forward characteristic. In view of the rapid increase in reverse current it is essential in practice to limit the reverse voltage that is applied to a unit. The maximum allowable figure is between 7 and 9 volts r.m.s., i.e. 10 to 12·7 volts peak. In order to withstand voltages in excess of these, a number of units must be placed in series.

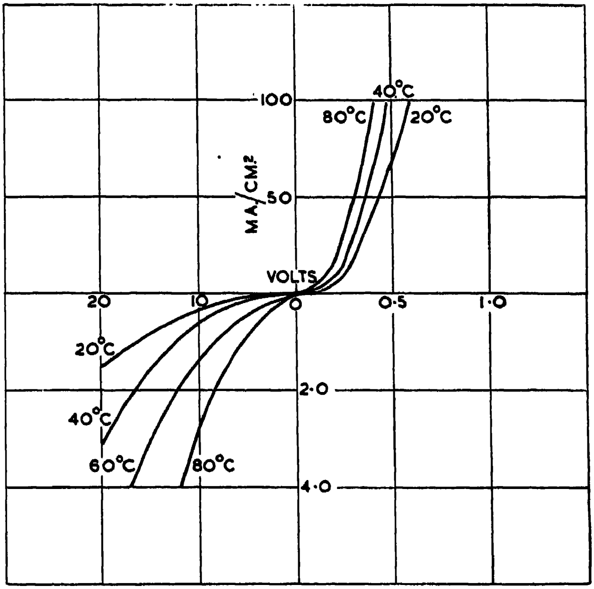


FIG. 13–2

CREEP

The current flowing in the forward direction of a rectifier depends only on the applied voltage and the rectifier temperature. In the reverse direction, in addition to depending on both these quantities, the current also depends on the time for which the voltage is applied. In cases where the current tends to increase with time the creep is termed "positive," and where the reverse effect occurs the creep is termed "negative." On first applying the voltage the rate of change of current is relatively high, thereafter decreasing to zero with the current settling to a value corresponding to the particular voltage, temperature, and rectifier condition.

In the case of the copper-oxide rectifier creep is generally positive. Creep characteristics may be either instantaneous or otherwise, i.e. the current may be recorded immediately after applying a reverse voltage or a period may be allowed to elapse until the current has become stable. In the latter case the current values are, of course, greater for the same voltage, than in the former. After the voltage is removed, the reverse resistance slowly returns to its original instantaneous value, the rate of return being relatively high initially and thereafter decreasing. Providing a sufficiently long period is allowed before reapplication of a reverse voltage, the creep characteristic is reproducible. If, however, a voltage is reapplied before the initial conditions are attained the characteristic will not be repeated. If a stable current value has been reached and a smaller voltage is then reapplied, creep will be negative; if a larger voltage, positive creep will result.

AGEING

The irreversible change, known as ageing, is caused by the effect of temperature with regard to the forward direction, and by temperature and reverse voltage with regard to the reverse direction. The effect of temperature is to increase both the forward and reverse resistances. At room temperature the rate of increase of resistance is very slow, but becomes extremely rapid at excessive temperatures. In practice, rectifier temperatures should normally not be allowed to exceed 55° C. The effect of reverse voltage is to reduce the reverse resistance and, in practice, this effect is of greater importance than that of temperature. Hence it is desirable to employ as low a reverse voltage per rectifying unit as is practicable.

The Selenium Rectifier

In many respects selenium rectifying units are very similar to those of copper-oxide. The unit consists of a metal support (usually a disc) on which is deposited a layer of selenium. The disc is either of aluminium or iron, generally the latter, and when iron is employed it is usually nickel-plated and roughened to ensure good adhesion of the selenium. Selenium exists in three forms known as vitreous, α, and β. The vitreous form is practically an insulator, the α form has a high specific resistance, while the β form has a somewhat lower resistance. It is the latter form which is employed in the production of rectifiers.

In order to apply the selenium to the metal support, the vitreous form is reduced to a powder, which is then evenly spread over the support. This is then heated to a temperature of, approximately, 125° C. while pressure is simultaneously applied. Under these conditions a homogeneous layer of uniform thickness is produced which changes to the α form on lowering the temperature to 100° C. The thickness of the layer is of the same order as that of the oxide in the copper-oxide rectifier, i.e. 0·1 mm. The selenium is next converted to the β form by heating it to approximately 200° C., the time of heating varying somewhat with the actual temperature. Following the above process, a counter-electrode consisting of a low melting-point alloy (such as lead, tin, and bismuth) is sprayed over the entire surface of the selenium, except for two narrow rings at the centre and edge of the disc. These rings are, of course, to prevent contact occurring between the counter-electrode and the support.

The theory of the selenium rectifier is very similar to that of copper-oxide; i.e. it depends for its action on the existence of a barrier layer situated between a good and a poor conductor. The latter consists of selenium of the β form which, by the addition of agents having a similar effect to an excess of oxygen in cuprous oxide, has a higher conductivity than normal. Thus, whereas the pure β form has a specific resistance of several thousand ohms, when employed in the rectifier the resistance is usually only a few hundred ohms.

In the case of the selenium rectifier it is found that the direction of electron flow is from the selenium to the metal base, which indicates that the barrier layer is not situated between these two. Actually the barrier layer is located between the counter electrode and the selenium. After the production process described above it is found that the rectifying unit possesses very little asymmetrical conductivity. In order to produce the necessary asymmetry an electrical forming process must be employed. This consists of applying a voltage to the unit in the reverse direction for several hours. Initially the voltage must be low in order to prevent overheating, but as the reverse resistance increases, the voltage is increased until a safe maximum is reached. A proof that the barrier layer is formed between the selenium and the counter-electrode is that if the latter is removed, and a fresh one applied, the asymmetry is destroyed, and the forming process must be repeated. The change in characteristics consequent upon forming a unit is shown by Fig. 13–3 for a typical case. It will be noted that both forward and reverse resistances are increased, but the latter to a much greater extent than the former. After a unit has been formed it remains stable, but there is some reduction in reverse resistance after long storage. Thus, if a reverse voltage is applied to a rectifier which has been out of action for some time it will be found that, initially, the reverse current is higher than normal. The reverse resistance, however, quickly reforms to its normal value.

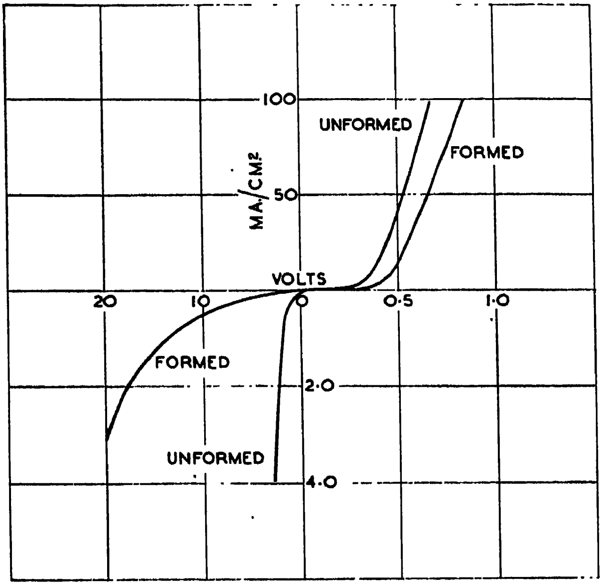


FIG. 13–3

ELECTRICAL CHARACTERISTICS OF SELENIUM RECTIFIERS

The characteristics of a selenium rectifying unit corresponding to those of the copper-oxide type are shown by Figs. 13–4 and 13–5. Referring to Fig. 13–4, it will be noted that the rectifier does not obey Ohm's Law, the resistance falling with an increase in current. From Fig. 13–5 it will be noted that the volt/ampere characteristics are very similar to those of copper-oxide, with a notable exception that the current turns sharply upwards at 0·5 volt instead of 0·25

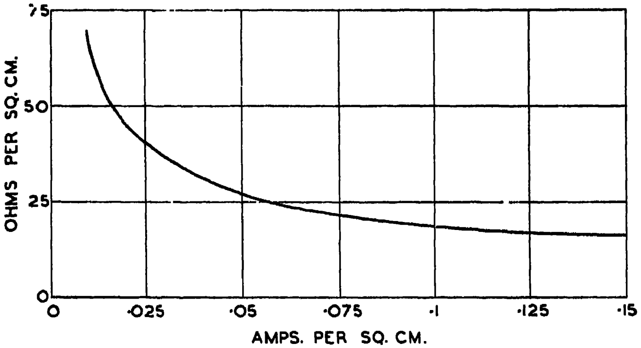


Fig. 13–4

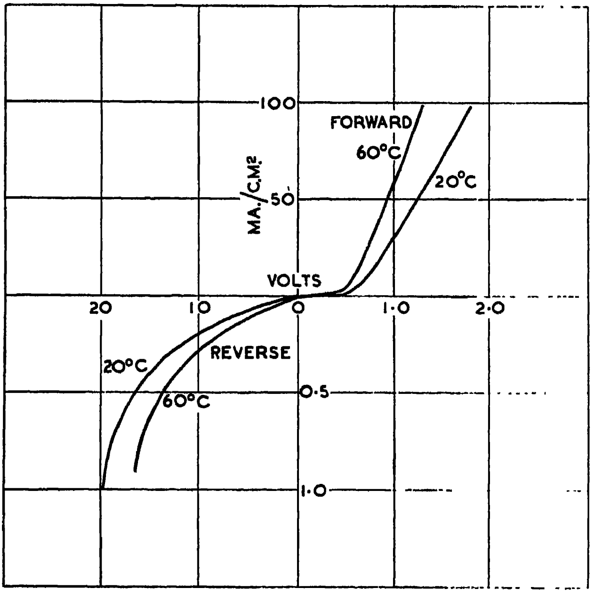


Fig. 13–5

volt as with copper-oxide. It is evident that the reverse resistance of the selenium rectifier is higher than that of the copper-oxide and, furthermore, less affected by temperature.

The Hot-cathode Rectifier

The thermionic gas-filled rectifier, or hot-cathode rectifier as it is commonly called, may be regarded as a development of the mercury-arc rectifier. From the viewpoint of power capacity it stands between the mercury-arc and metal rectifiers, and was introduced at about the same time as the latter. It consists of a heated oxide cathode and two or more nickel or graphite anodes enclosed in a glass envelope with an argon or mercury-vapour filling. Some typical examples are shown by Figs. 13–6 and 13–7. That of Fig. 13–6 is a B.T.H. BD12 type with a maximum mean current output of 33 amp., the maximum permissible voltage between anodes being 2 × 125. The gas-filling is mercury vapour. That shown by Fig. 13–7 is a Philips 367 type, the maximum mean current being 6 amp. and the maximum voltage between anodes 2 × 45. The gas-filling is argon. The laws governing the behaviour of hot-cathode rectifiers are substantially the same as those of the thermionic gas-filled diode treated in Chapter IX. However, due to the presence of two or more anodes in the same tube, further limitations occur.

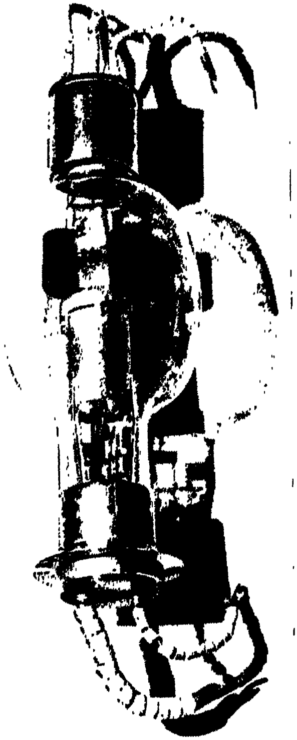


FIG. 13–6

By courtesy of the B.T.H. Co., Ltd.

Cross-firing

When two anodes, for the purpose of full-wave rectification, are enclosed in the same tube, a possibility of a direct discharge between anodes, as well as back-firing, occurs. In the case of a bi-phase rectifier the maximum potential difference between anodes is approximately the same as that between anode and cathode, i.e. twice the phase to neutral voltage. Hence the practice has arisen of specifying the maximum permissible voltage between anodes as 2 × V where V is customarily expressed in r.m.s. values. Thus, for the two valves referred to above, the maximum phase to neutral voltages are, respectively, 125 volts and 45 volts r.m.s. If the voltage between anodes is gradually increased, a figure will be reached at which a discharge will occur. The discharge will, initially, be of the glow type, because normally the anodes are "cold." The voltage at which striking between anodes takes place depends on the gas-filling, its density, and the spacing of the anodes. With regard to density and spacing, the striking potential follows Paschen's Law and, for the usual gas-filling pressures, falls with an increase of density and pressure. Hence, in mercury-filled tubes, the cross-firing voltage will depend on temperature in a similar manner as does the maximum inverse voltage.

Prior to a glow discharge occurring between anodes, or between anode and cathode when an anode is negative, a small current may be observed to flow to the "non-conducting" anode, providing a current is flowing to the conducting one in the normal manner. While the latter current exists so also does a plasma, and the anode which is at a negative potential will draw a positive-ion current from the plasma in the same manner as does a probe or the grid of a thyratron. As the negative potential of the "non-conducting" anode is increased a stage will be reached at which normal breakdown occurs with a glow discharge. Hence at this point the "non-conducting" anode current will tend to increase rapidly. In order to indicate the nature and magnitude of these phenomena, experiments were made by the author on a small argon-filled valve. The results of Fig. 13–8 were obtained by means of the arrangement shown by Fig. 13–9. Curve (1) shows the inverse breakdown voltage and glow current under open-circuit conditions. It will be noted that the maximum inverse voltage is 160 volts. On causing a current of 1 amp. to flow to the conducting anode, Curve (2) was

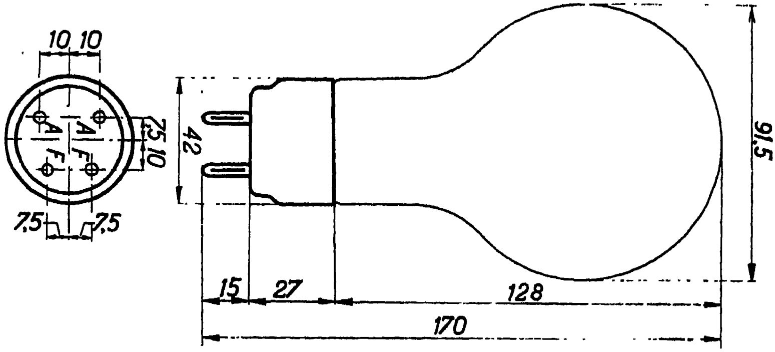


FIG. 13–7*

---

* All dimensions are in mms.

obtained. It is evident that at a few volts negative a positive-ion saturation current occurs very similar to that of a thyratron grid or probe. As the voltage is increased, the current increases slowly at first and then much more rapidly as the normal breakdown voltage is approached. The absence of a sharp increase at the breakdown voltage is probably because the positive-ion current which precedes the glow discharge is considerably larger than the initial value of the latter current. On increasing the conducting-anode current to 2 and 3 amp., Curves (3) and (4) were obtained. The increased values of the currents of these are, of course, due to the increase of

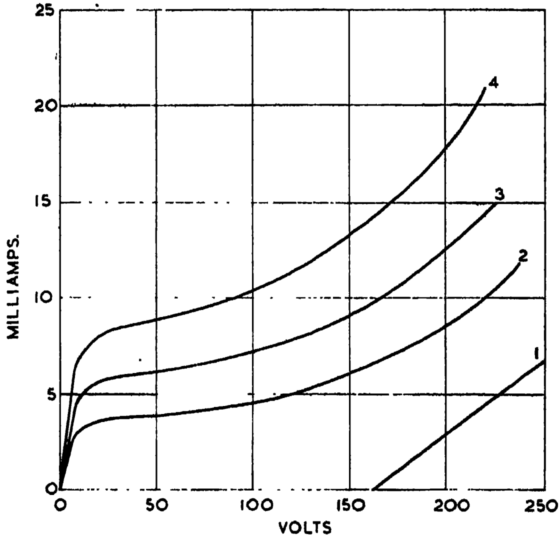


FIG. 13-8

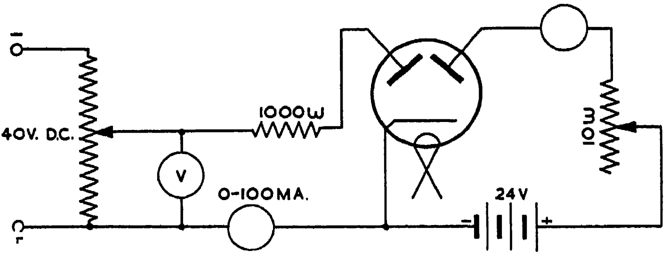


FIG. 13-9

positive ions within the plasma. It is apparent that under all conditions the "non-conducting" anode current is an increasing function of the anode-cathode voltage and hence the glow is "abnormal." This, of course, indicates the likelihood of the glow passing over to an arc with a complete failure of the valve.

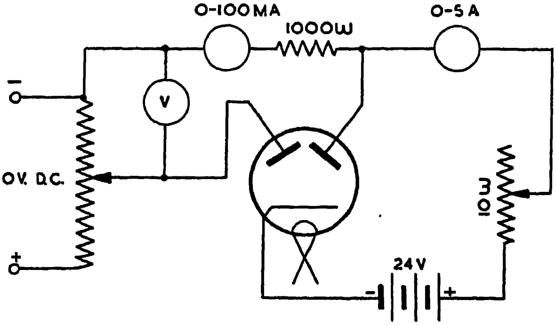


FIG. 13–10

The arrangement of Fig. 13–9 was adopted mainly to show the probe-like character of a non-conducting anode. Practical conditions are more nearly represented by Fig. 13–10 and the curves of Fig. 13–11 for a different, but similarly rated, valve were obtained

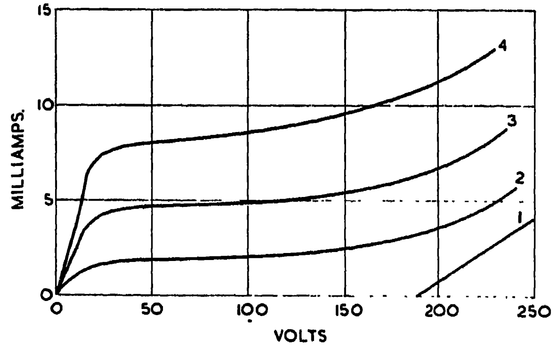


FIG. 13–11

with the latter arrangement. It will be noted in both cases that the positive-ion current is roughly proportional to the conducting-anode current. This, of course, should be so for, as we have already seen when dealing with probes, the positive-ion current is proportional to the positive-ion density in the plasma. The curves show that beyond a certain anode voltage the anode-to-anode current rapidly increases, and also that this current tends to be proportional to the load current. Hence cross-firing is more likely to occur at heavy loads than otherwise. It is evident that a cross or back-fire is least likely under no-load conditions.

In order to indicate the conditions just prior to a back-fire in the valve of Fig. 13–10, the anode-to-anode voltage was considerably increased and the results shown by Fig. 13–12 obtained. It was found that the tendency to back-fire became distinct at about 50 mA, the possibility of the current rising to a dangerous value beyond this figure being quite evident. The mean anode current of this valve being 3 amp., it appears desirable from Fig. 13–12 to limit the maximum voltage between anodes to that at which the rate of increase of glow current becomes pronounced, i.e. 2 × 75 volts peak or 2 × 50 volts r.m.s. approximately. Apart from the danger of a back-fire, a glow discharge leads to losses the magnitude of which may be of a serious nature. This is sufficiently evident from Fig. 13–12.

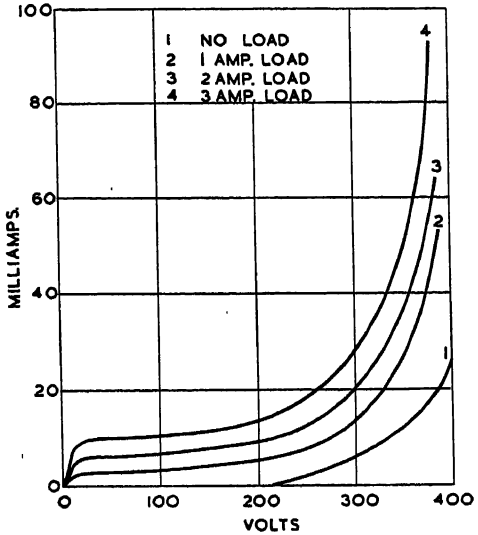
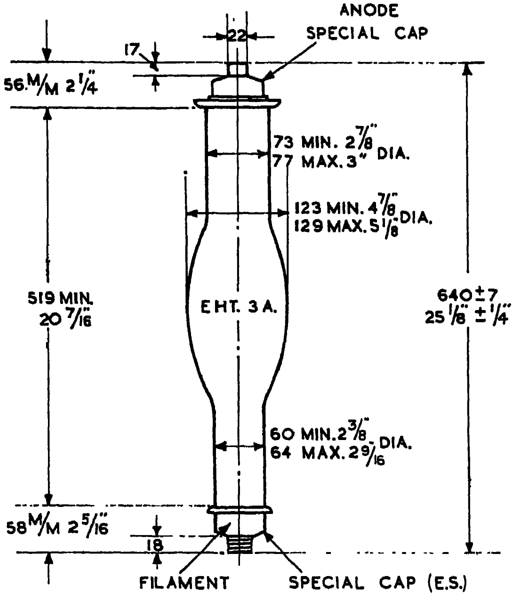


FIG. 13–12

High-vacuum Rectifying Valves

High-vacuum valves are employed for rectifying purposes whenever large inverse voltages must be withstood and where the reverse current must be negligible. Thus such valves are largely employed in producing the d.c. supply for X-ray tubes and for high-tension cable testing. In small sizes vacuum valves are frequently employed for the d.c. supply in wireless receivers. In such cases a full-wave valve is customary, i.e. two anodes are enclosed in the same envelope.



ALL DIMENSIONS ARE IN M/M
AND ARE MAX EXCEPT WHERE
OTHERWISE STATED
NET WEIGHT 2¾LBS

FIG. 13–13

By courtesy of the G.E.C., Ltd.

With this arrangement, however, difficulties exist in causing each anode to collect efficiently the electrons from a common cathode. Hence each anode is usually provided with its own cathode, the cathodes being connected either in parallel or series.

For high-voltage work diodes of the type described in Chapter VIII are employed. This is principally because of the difficulty of bringing out two anode leads in a manner to avoid flash-over.

Some dimensional details of a EHT3A valve are shown by Fig. 13–13, this valve being capable of withstanding a maximum inverse

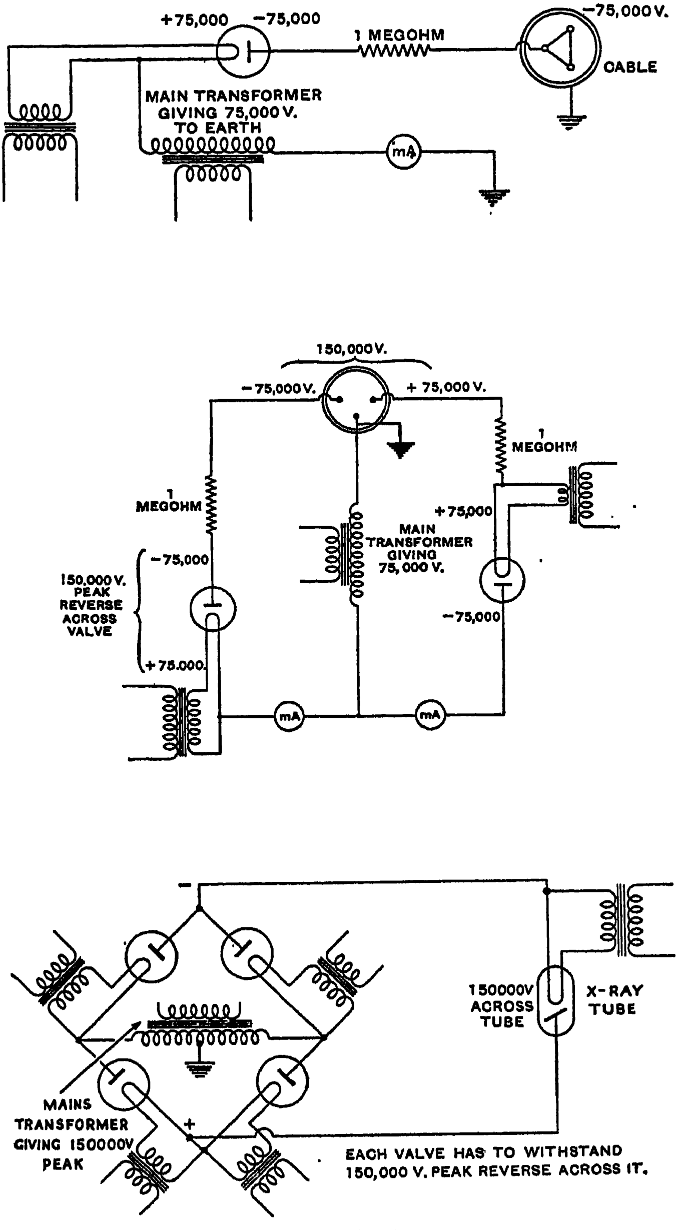

+75,000
−75,000
1 MEGOHM
−75,000 V.
CABLE
MAIN TRANSFORMER GIVING 75,000 V. TO EARTH
mA
150,000 V.
− 75,000 V.
+ 75,000 V.
1 MEGOHM
1 MEGOHM
+75,000
MAIN TRANSFORMER GIVING 75,000 V.
−75,000
150,000 V. PEAK REVERSE ACROSS VALVE
−75,000
+75.000
mA
mA
−
150000V ACROSS TUBE
X-RAY TUBE
MAINS TRANSFORMER GIVING 150000V PEAK
+
EACH VALVE HAS TO WITHSTAND 150,000 V. PEAK REVERSE ACROSS IT.

Fig. 13–14

By courtesy of the G.E.C.. Ltd.

TABLE 13–1

CHARACTERISTICS

Filament voltage	13·0 max.
Filament current	9·0 amp.
Max. reverse peak voltage	150 kV
Peak (for max.) rectified current (total emission)	200–250 mA
Max. mean rectifier current for single-valve circuit	83 mA
Max. mean rectifier current in Gratz full-wave circuit (4 valves)	166 mA
Impedance	750 ohms approx.

Cap: Edison Screw for Filament.
Anode Cap.

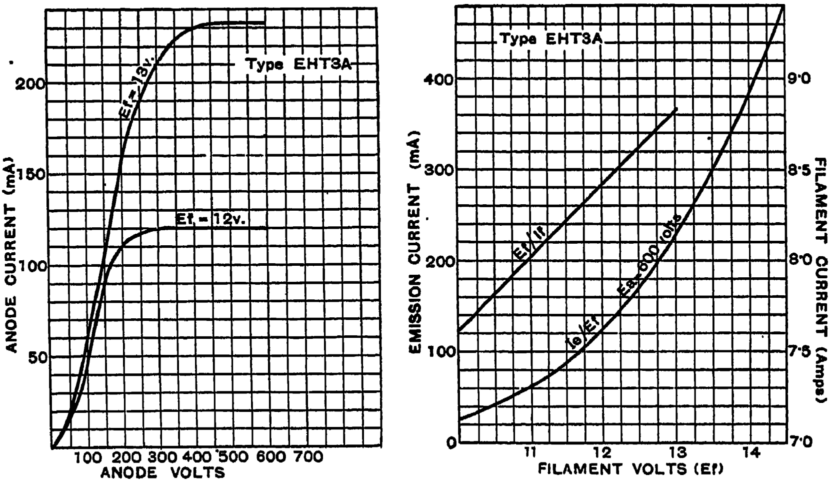


FIG. 13–15

By courtesy of the G.E.C., Ltd.

voltage of 150,000 volts. Table 13–1 gives some characteristics of the valve and Fig. 13–14 typical circuits in which it may be employed. Emission current curves are shown by Fig. 13–15.

The Mercury-arc Rectifier

The mercury-arc rectifier principally differs from the hot-cathode mercury-vapour type in that the cathode takes the form of a pool of mercury from which electrons are liberated by one of the methods previously described in Chapter II. Constructionally, three forms of mercury-arc rectifiers may be distinguished: the glass-bulb type, shown by Figs. 13–16 and 13–17; the steel-tank variety, illustrated by Fig. 13–18; and an intermediate form, known as the steel-bulb

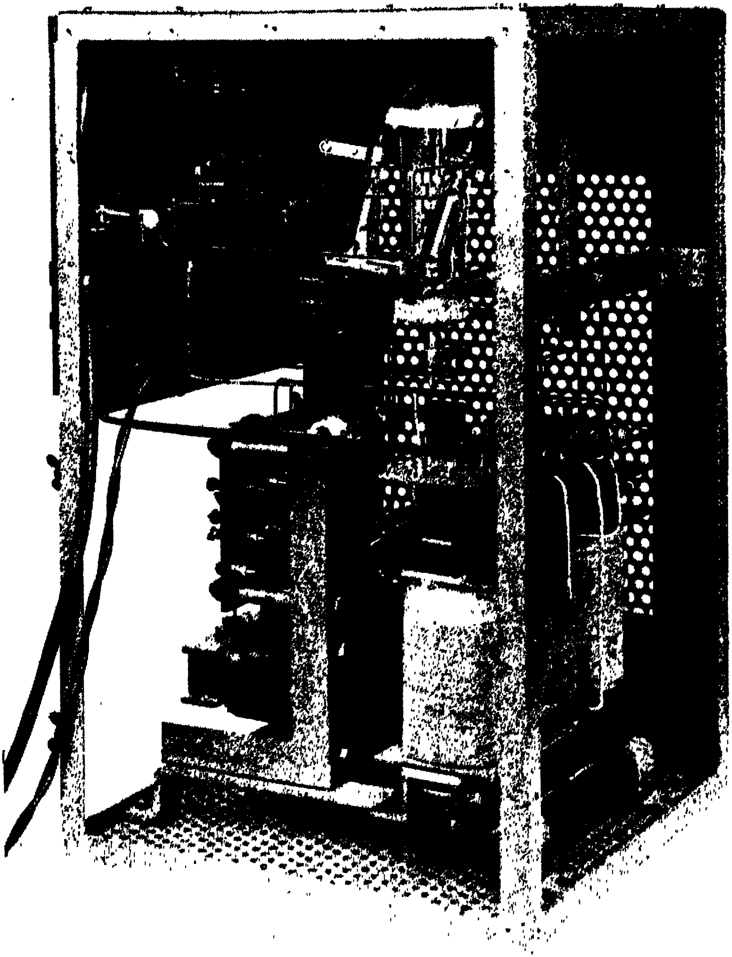


FIG. 13-16

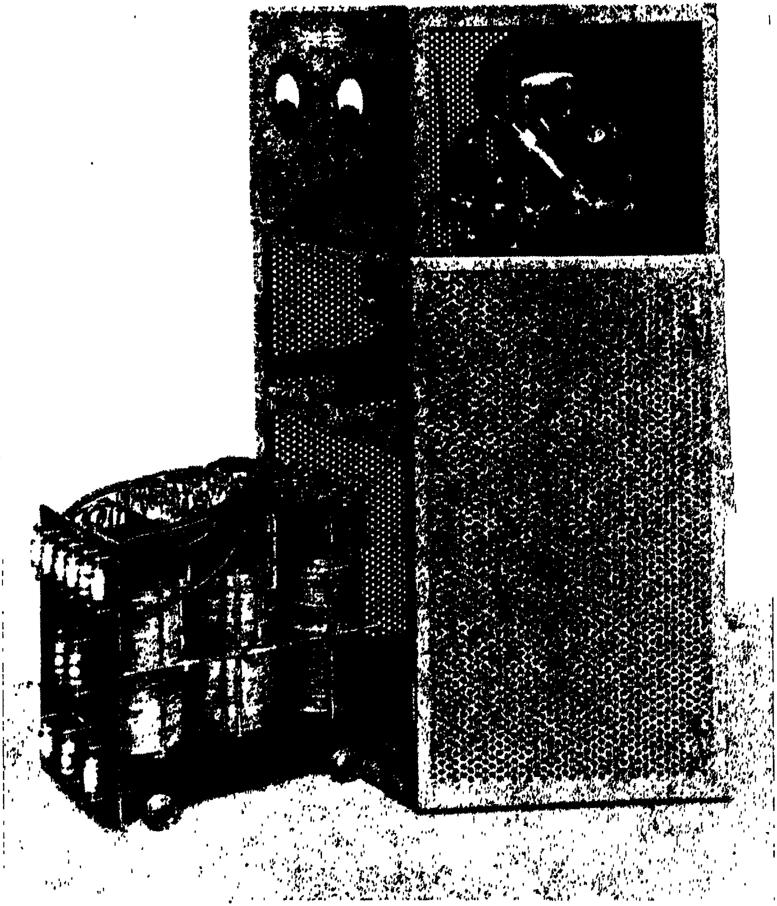


FIG. 13–17

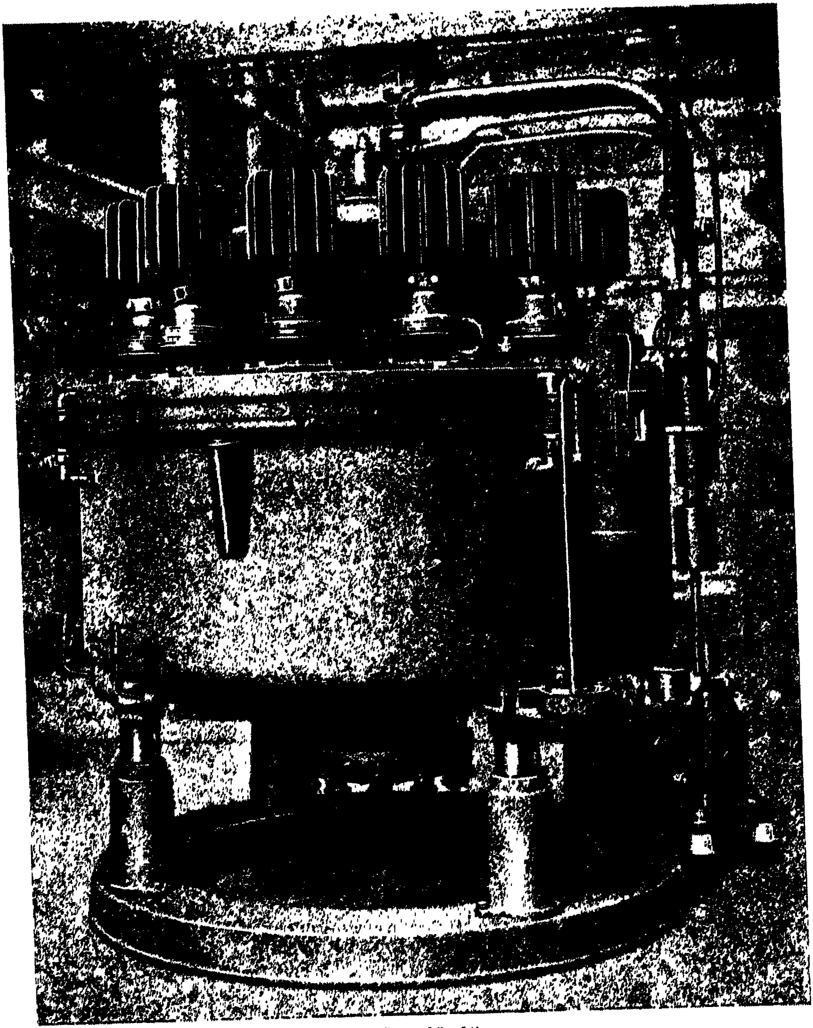


Fig. 13-18

By courtesy of the G.E.C., Ltd.

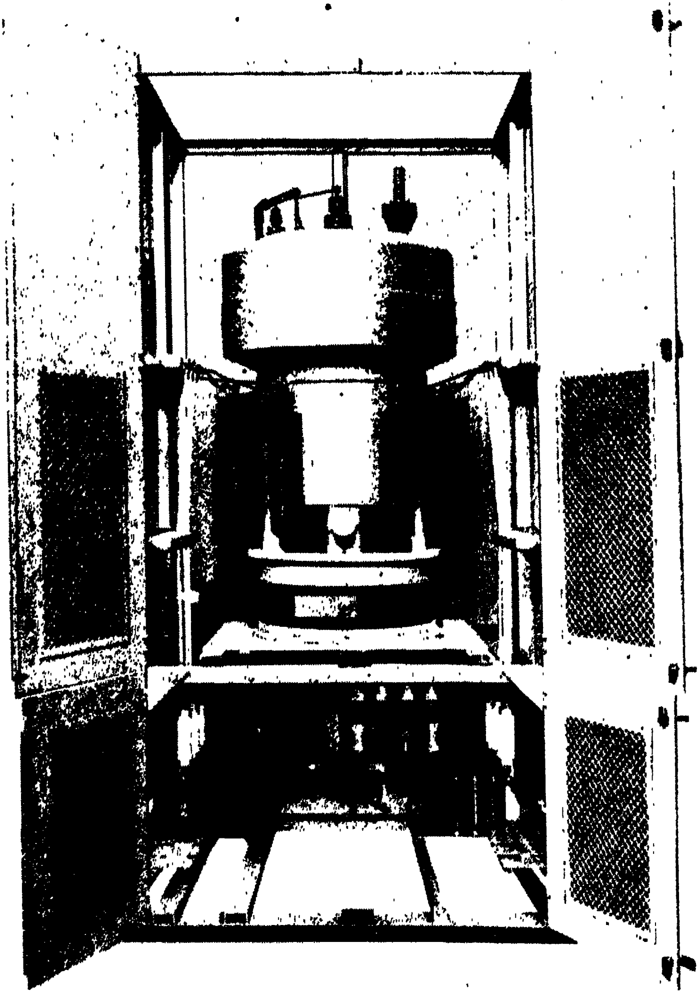


Fig. 13–19

By courtesy of the G.E.C., Ltd.

type, shown by Fig. 13–19. The first and last are sealed off and have a maximum current capacity of about 500 amp., with a limiting power capacity of 250 kW. The steel-tank rectifier is housed in a vacuum chamber of steel, is demountable, continuously evacuated, and water-cooled. In this form rectifiers have been constructed for as much as 16,000 amp. at relatively low voltages, 2500 kW at 3300 volts, and voltages as high as 20 kV at 600 kW. With regard to an upper voltage limit, voltages as high as 30 kV may be obtained, but only by derating the current output as shown by Fig. 13–20. The reason for this will appear later. Unlike the hot-cathode rectifier, the mercury arc has no filament losses or deterioration, and the capacity for overload is not limited by

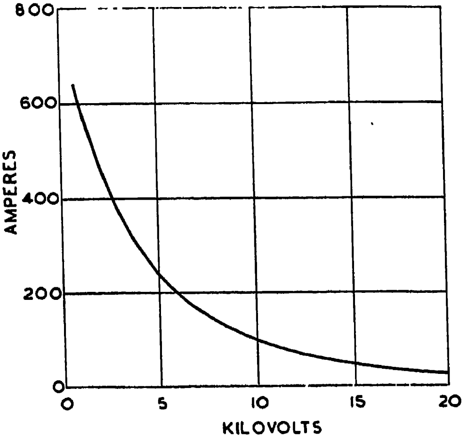


Fig. 13–20

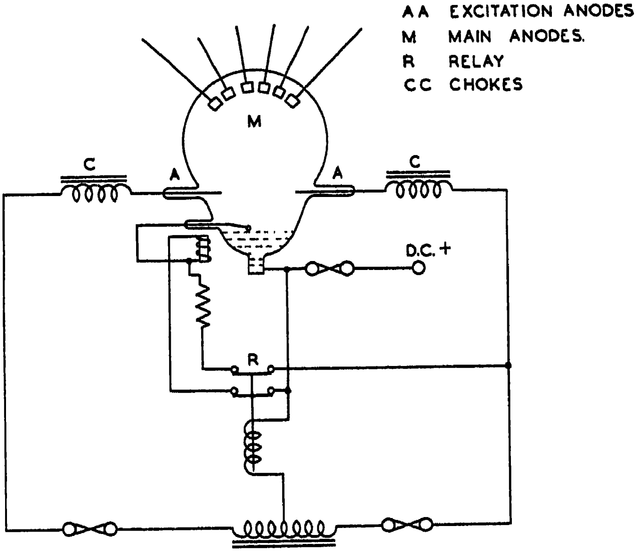


Fig. 13–21

cathode conditions. Against this must be set the necessity of mechanism for striking and maintaining the arc, as indicated by Fig. 13–21.

THE GLASS-BULB RECTIFIER

A typical glass-bulb mercury-arc rectifier with the necessary mechanism for striking and maintaining the arc is shown by Fig. 13–21. The arc is struck by the movable-ignition electrode, which is drawn by the electromagnet into the cathode pool and then released while current is flowing. This strikes an arc which, on opening of the ignition circuit by the ignition relay, is transferred to the excitation anodes. This holding-arc, as it is sometimes termed, is essential, for, in its absence, should the main load and arc be reduced below a certain value, extinction occurs necessitating restriking of the arc. The function of the chokes is to produce overlapping of the individual excitation anode currents, so that at no moment is the excitation current zero.

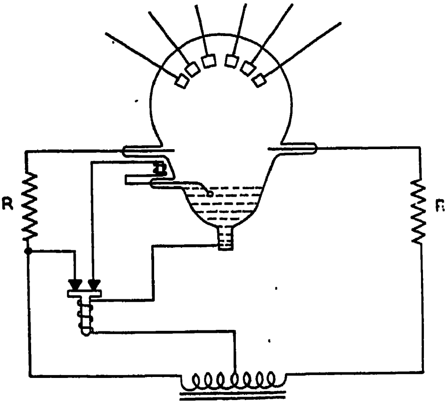


FIG. 13–22

An alternative but similar method of ignition and excitation to that described above is indicated by Fig. 13–22. Here it will be noted that ballast resistances *RR* replace the chokes of Fig. 13–21 and that the ignition electrode initially rests in the cathode pool. On closing the supply switch the electromagnet withdraws the ignition electrode from the pool, thereby striking an arc which permits the establishment of the excitation arc. The rectified current due to the latter then flows through the ignition relay, which in consequence opens and interrupts the ignition circuit. The necessary overlap of the excitation currents is effected by the self-inductance of the relay winding.* Compared with that of Fig. 13–21, an advantage of Fig. 13–22 appears to be a somewhat simpler relay.

During operation of the rectifier, the body of the bulb is filled with ionized mercury vapour, emitted at high velocity from the cathode pool, from which it proceeds to the condensing surfaces.

---

* In some motor applications the choke is replaced by the motor shunt field. See "A Variable-Speed Power Drive," *The Engineer*, Nov. 8th, 1946, p. 423.

The evaporating mercury has a velocity of the order of 20 kilometres per second, and to avoid this vapour blast the anodes are placed in glass arms as shown. From where the vapour is emitted is known as the "cathode spot," this being initially formed in striking the arc. Evaporation of the cathode is a natural accompaniment of arc conduction and it follows that for continuous operation the only practicable cathode is one which is a liquid at normal temperature. In this manner the cathode is self-restoring as the metal vapour condenses on the sides of the containing chamber.

The cathode spot emits a brilliant light and acts as a copious source of free electrons. On the thermionic theory these are produced by reason of the spot temperature, the latter being maintained by positive mercury ion bombardment. The high-velocity vapour blast deflects these ions, which are, in consequence, ever seeking fresh points of contact on the cathode. The result is that the spot travels over the mercury pool in a highly erratic manner.

The current density of the spot is about 4000 amp. per cm.^{2} and its temperature has been variously estimated at from 600° C. to 3000° C. Determination of the spot temperature by direct-contact methods gives results of the order of the lower figure quoted, while pyrometric measurement indicates the temperature to be equivalent to that of a black body at 2400 K. The working temperature of the cathode lies between 120° C. and 200° C.

In some experiments carried out by Teago and Gill* the cathode spot was confined to a small area and a platinum-iridium couple arranged to be in the hottest region. With a current of 4 amp., the recorded temperature was 370° C., while a current of 15 amp. produced a temperature of 750° C.

THE ARC VOLTAGE DROP

Unlike the hot-cathode rectifier, the cathode drop in the mercury arc is not lower than about 10 volts. This is due to the positive ions having either to fall through a sufficient potential to liberate electrons from the cathode, or to produce a sufficient gradient to extract them by field emission. Such processes are, of course, unnecessary when a separately heated cathode is employed. The total anode-cathode drop consists of three parts, that already mentioned, the drop in the arc, and the drop at the anode. The latter may vary from 2 to 8 volts, while variations in the arc drop are from 8 to 15 volts. The cathode drop remains approximately constant, the arc and anode drops varying with the loading, vapour pressure, and design of the rectifier. Fig. 13–23 for the rectifier of Fig. 13–24 shows the anode-cathode voltage drop for various current loadings. It will be noted that the arc strikes at about 30 volts, the drop then rapidly decreasing with increasing current until an approximately constant figure of about 19 volts is reached. Although for practical and design purposes the anode-cathode drop may be regarded as constant, it is evident that it possesses a minimum, rising beyond a certain current value. The initial fall of Fig. 13–23 is due to the increasing conductivity of the arc with increasing current. The rise in voltage, as the current is increased beyond the value at the minimum, is due to the increasing current density in the arc, the latter being unable to expand because of the constricting effect of the anode arms or shields. As stated in Chapter II, the voltage drop in a gaseous discharge is a function of the gas density and hence the mercury-arc drop depends on the vapour density and thus on the vapour temperature. Immediately on switching on a mercury-arc rectifier from cold, it is found that the output voltage tends to increase during the first few minutes of operation. This is due to the rise in temperature and mercury-vapour pressure, the relation between these quantities being shown by Table 13–2.

---

* *Mercury Arcs* (Methuen), p. 6.

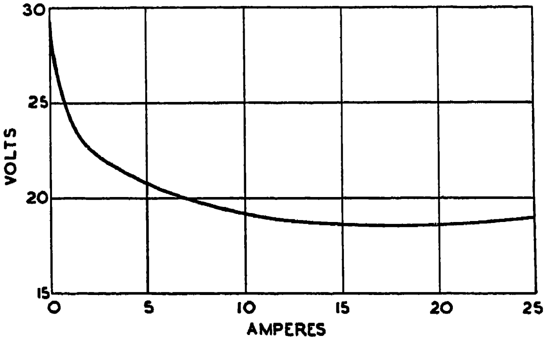


Fig. 13–23

In the cold state the vapour density is such that the rate of ion formation is lower than at high temperature, with the result that the conductivity of the arc tends to be low with a relatively high drop for a given current. As the vapour density increases, so also does ionization, resulting in a reduced arc drop. Fig. 13–25 shows the rise in output voltage of the rectifier of Fig. 13–24 from the moment of starting from room temperature at full load.

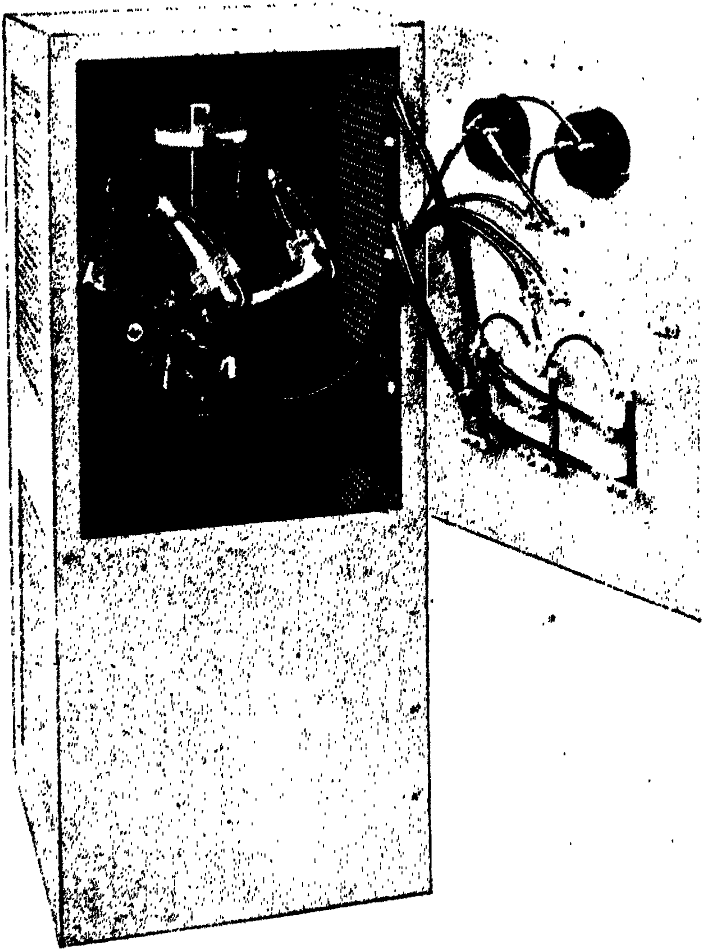


FIG. 13–24

TABLE 13–2

	Temperature ° C.	Saturated Vapour Pressure mm. of Hg
	0	0·00016
	5	0·00026
	10	0·00043
Normal Working of Mercury arc Rectifier	15	0·00069
	20	0·00109
	25	0·00168
	30	0·00257
	35	0·00387
	40	0·00574
	50	0·0122
	60	0·0246
	80	0·0885
	100	0·276

The anodes of the mercury-arc rectifier are formed of either graphite or iron, usually the former for the glass-bulb type and the latter for the steel-tank rectifier. The purpose of the employment of these materials is that neither are wetted by mercury and both have poor electron-emitting properties. As previously shown, the anodes are housed in bent-glass arms in the glass-bulb rectifier and, to increase their heat-radiating properties, the anodes are usually fluted. Besides protecting the anodes from the vapour blast, the glass arms are also effective in preventing cross- and back-firing.

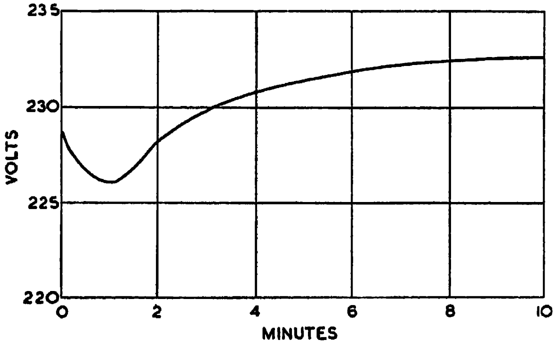


Fig. 13–25

Surmounting the pool and arms of the glass-bulb rectifier is the condensation chamber shown by Fig. 13–17. The function of this is, of course, to assist in condensing the mercury vapour and, by virtue of its heat-radiating area, maintain the rectifier unit at the required temperature.

The Steel-tank Rectifier

On the assumption that the anode-cathode voltage drop is constant, it follows that the arc losses are proportional to the d.c. output current. Thus, for a given temperature rise the radiating surface of a mercury-arc rectifier must be proportional to the current output. It is, of course, this fact which limits the size of the glass-bulb rectifier and for currents in excess of about 500 amp. the water-cooled steel-tank rectifier is employed.

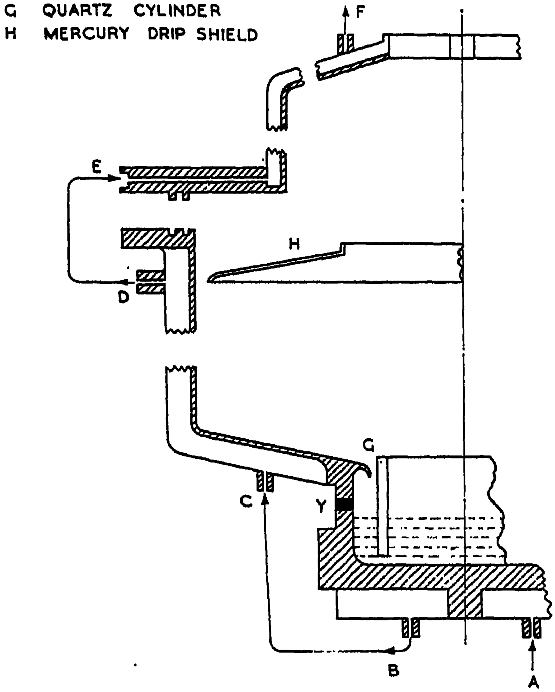


Fig. 13–26

A typical rectifier of the latter type is shown by Fig. 13–26. It will be noted that the large condensing dome, prominent in the glass-bulb type, is absent as, of course, cooling is effected by water circulation. The anodes are attached to the top plate and are shielded by means of metal or insulated shields or "sleeves." The mercury pool is insulated from the tank, for otherwise the arc may transfer to this, destroying the steel where the cathode spot forms. The cathode spot is confined to the central part of the mercury pool by means of a quartz cylinder, and it will be noted that the bottom casting of the rectifier overhangs the pool in order that the condensed mercury in returning does not short-circuit the cathode insulation *Y*.

COOLING

Cooling of the rectifier is effected by water first passing in at *A* and then round the mercury pool. It leaves at *B* by means of a rubber pipe, and re-enters the bottom casting at *C*. A rubber or other insulating pipe is, of course, essential to avoid short-circuiting the cathode insulation. The water now leaves the bottom casting at *D* and enters the top casting radially at *E* by means of a metal pipe, finally leaving at *F*.

In view of the fact that the temperature of the rectifier is chiefly limited by water circulation it is unnecessary to make the arc chamber any larger than is necessary to accommodate the anodes within it. Thus a 600-volt 1500-kW rectifier has a container about 6 ft. in diameter and 6 ft. high.

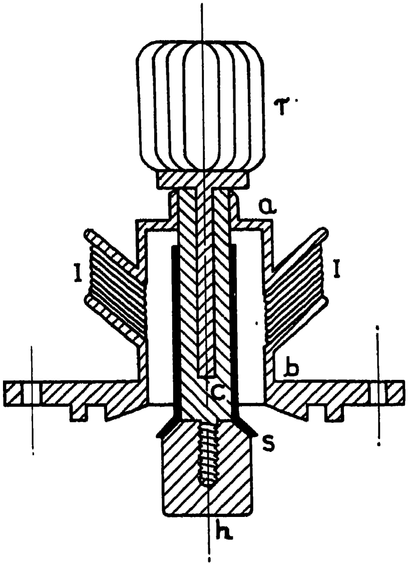


FIG. 13-27

VACUUM TECHNIQUE

A highly important feature of the steel-tank rectifier is, of course, the means whereby the vacuum is effected and preserved. The anode sealing is very important because the anode stems must be insulated where they pass through the anode plate. At this point both the potential difference and the temperature are a maximum.

THE VITREOUS SEAL

A typical anode assembly and seal of the vitreous type is shown by Fig. 13–27. The anode body consists of two steel castings separated by an insulator *I*. The insulator is formed by a number of thin mild-steel cones which are separately enamelled with a special glass having the same coefficient of expansion as mild steel. After assembly the cones are fused together and to *a* and *b* in an electrically heated oven, the result being a gas-tight joint of high dielectric strength. To the top casting is welded the anode stem *c*, which carries the renewable head *h*. Surrounding the anode stem is an insulating sleeve *s*, which serves to concentrate the discharge on to the anode head. To the top of the anode is attached the radiator *r*.

This form of anode assembly is fitted to the anode plate by means of a tongue and groove joint as shown. The joint is rendered gas-tight by means of a lead gasket.

The Mercury Seal

The arrangement of a mercury anode seal is shown by Fig. 13–28. The porcelain anode insulator is located by spring pressure on an asbestos ring *a* in a circular recess in the anode plate. This recess is sealed off from the atmosphere by a layer of mercury which is sealed at its upper end by the rubber seal *r* held by the gland *g*. The mercury is poured into the gauge-glass from which the permanency or otherwise of the sealing may be observed. This seal has the property of being able to take up differential expansions and contractions with a minimum of stressing.

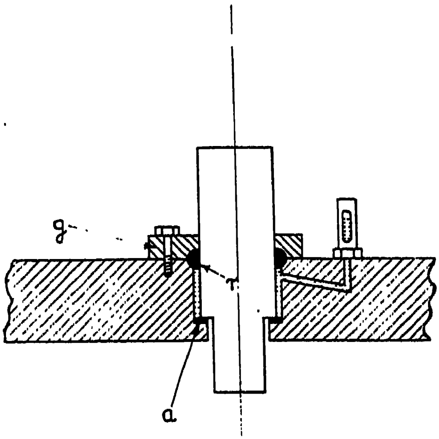


Fig. 13–28

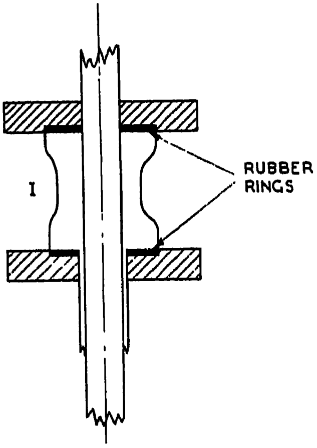


Fig. 13–29

The Rubber Seal

The rubber seal is shown by Fig. 13–29. It consists of a wide ring of rubber reinforced on the vacuum side by a thin flexible iron ring. The purpose of the latter is to prevent gas from the rubber entering the arc chamber and also to prevent mercury vapour from contaminating the rubber. It will be noted that the rubber rings are located in recesses in the anode plate and top of the anode stem, the latter passing through a large porcelain insulator *I*.

VACUUM MAINTENANCE

No matter how good the anode seals may be, or how small air leakage is, the vacuum of the steel-tank rectifier cannot be continuously maintained at the necessarily high figure essential to satisfactory operation without permanently attached pumping equipment. Apart from leakage, this is due to the release of occluded gases from the walls of the container during operation of the rectifier. The pumping system consists of an oil or mercury-vapour diffusion pump backed by a rotary-vane pump operating under oil. The latter produces the preliminary vacuum, while the vapour pump produces the high vacuum necessary for satisfactory working.*

IGNITION AND EXCITATION OF STEEL-TANK RECTIFIERS

In the steel-tank rectifier ignition and excitation may be effected by a single-phase system similar to those employed with the glass-bulb rectifier and shown by Figs. 13–21 and 13–22. For several reasons, however, such as the elimination of the excitation anodes and to obtain greater stability under adverse conditions, a direct-current system of excitation is often employed. A typical system is shown by Fig. 13–30, in which it will be noted that a separate rectifier is employed. This is usually of the copper-oxide or selenium

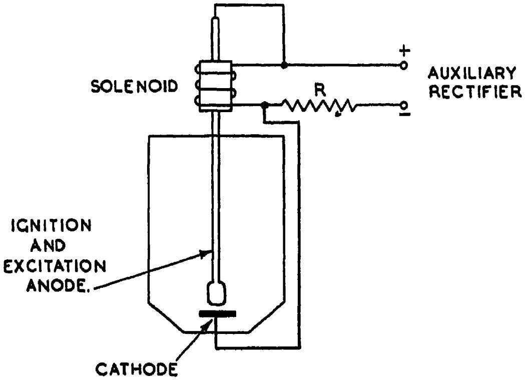


FIG. 13–30

---

* "Aircraft Instrument Pumps," F. G. Spreadbury, *Aircraft Production*, Aug. 1941, p. 296.

"The Theory of the Mercury-Vapour Vacuum Pump," P. Alexander, *J. Sci. Inst.*, Jan, 1946, p. 11.

types, preferably three or six phase. It will be further noted that there is only one auxiliary anode, this serving both for ignition and excitation purposes. On closing the supply switch, current from the auxiliary rectifier passes through the resistance R and ignition solenoid. The action of the solenoid causes the ignition electrode to dip into the mercury pool, thereby short-circuiting the solenoid. This releases the ignition electrode, whereupon an arc strikes which is maintained by current from the auxiliary rectifier. The volt drop across the solenoid is now equal to that across the arc, and this is less than that before the arc was struck and insufficient to cause a further descent of the ignition rod. In the event of a failure of the supply and its subsequent restoration, automatic re-striking will, of course, occur.

It will be observed that a further advantage of the system of Fig. 13–30 is the elimination of the relay of Fig. 13–21.

The Ignitron

The ignitron is a form of mercury-arc rectifier in which the arc is struck in a different manner from that previously described. Striking is effected by a rod of semi-conducting material (*permanently* immersed in the mercury pool) through which a current is passed into the mercury. The material is boron or silicon carbide, and the current needed is several amperes under the influence of 100 volts or so. Ignition occurs within a few microseconds of the current passing. The theory of the igniter, as the rod is termed, seems somewhat uncertain, but Slepian and Ludwig have considered it as the case of a homogeneous semi-conducting rod dipped into a good conductor. They have shown that with these conditions potential gradients of the order of 10^6 volts per cm. might be made to exist at the surface of a mercury pool if a suitable current were forced through the igniter rod into the mercury. In these circumstances a movement of the mercury may occur, causing a breakage of the circuit between some of the particles of the rod and the mercury. It has been further suggested that circuit breakage occurs due to heating at the contact of igniter and mercury, thus causing a motion of the latter. Experiments with starting currents show that the magnitude of the current depends on the surface condition of the igniter rod and whether it is wetted by the mercury. If the latter effect occurs, the starting current is relatively high, as it is when the rod has a polished surface rather than a dull matt one. This indicates that it is desirable to have a poor contact between rod and mercury, which leads to the idea that the igniter should consist of many fine particles, which make contact with the mercury through their edges and points only.

It has been assumed by Cage* that ignition is due to the current flowing into the mercury through a narrow band of particles at a depth l beneath the surface of the mercury. Considering Fig. 13–31 for a given current in one particle in the band l, the current through this band is

$$I_1 = ad$$

where a is a constant and d is the diameter of the igniter rod. The current which descends below the band and thus enters the mercury beneath is

$$I_2 = bd^2$$

where b is another constant. Thus the total current is

$$I = I_1 + I_2 = ad + bd^2$$

Hence the ignition current should be a quadratic function of the igniter-rod diameter, and experiment shows that this is the case.

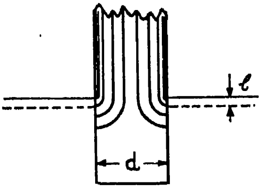


FIG. 13–31

It is evident that if at the moment at which ignition occurs the anode is at an adequate positive potential, an arc will occur and the ignitron will conduct. The anode-cathode potential will then, of course, adjust itself to the customary value for mercury-arc rectifiers, i.e. about 15 to 20 volts. In order that conduction may continue it is apparent that facilities must be provided for initiating the arc at each anode at the appropriate moment. Ignitrons are made with single anodes only, so that each constitutes a half-wave rectifier. For six-phase rectification six, then, are employed with all the mercury pools connected. If the pools are left unconnected, the fact that an ignitron fundamentally consists of half-wave rectifiers makes possible the use of the circuit of Fig. 14–10, if desired. It further follows that cross-firing is eliminated and that a holding-arc is unnecessary. It is claimed that the absence of the latter very much reduces the liability to back-fire.

Rectifier Losses

In the case of mercury-arc and hot-cathode rectifiers the valve-volt drop is, approximately, independent of the current and constant. Hence the rectifier losses are equal to the voltage drop multiplied by the mean value of the output current. For vacuum-type rectifiers the voltage drop across the valve is related to the anode current by

---

* "Theory of the Immersion Mercury-arc Igniter," J. M. Cage, *G.E.C. Review*, Vol. 38, p. 464.

$$i = Ae^{3/2}$$

and

$$e = A_1 i^{2/3}$$

Hence the instantaneous power loss is

$$ei = Ai^{5/3}$$

and the total loss

$$W = \frac{1}{t}\int_0^t iedt = \frac{A}{t}\int_0^t i^{5/3}\,dt \quad . \quad . \quad (13\text{–}3)$$

where t is the time for which each anode is conducting per cycle. Of course, to integrate (13–3), i must be expressed as a function of t.

METAL-RECTIFIER LOSSES

Referring to Figs. 13–2 and 13–5 it is evident that the theoretical determination of metal-rectifier losses is a matter of some difficulty. Not only is the voltage a function of the current, but there exist the complicating factors of temperature and creep. However, ignoring for the moment the two latter, we may, as an approximation, take the voltage/current relation in the forward direction to be

$$e = a + bi \quad . \quad . \quad . \quad (13\text{–}4)$$

where a and b are constants. As i and e are the instantaneous values of the current and voltage, the instantaneous power loss is

$$ei = ai + bi^2$$

Assuming the current wave to be sinusoidal, then

$$w = aI_m \sin\theta + bI_m^2 \sin^2\theta$$

where w is the instantaneous loss and I_m the maximum value of the current. The mean loss over a cycle is

$$W = \frac{m}{2\pi}\int_{\frac{m-2}{2m}\pi}^{\frac{m+2}{2m}\pi} (aI_m \sin\theta + bI_m^2 \sin^2\theta)d\theta$$

where m is the number of rectifier phases. Now the integral of the first term is a times the d.c. output current (see 13–8), while the integral of the second term is mb times the square of the r.m.s.

value of the current per phase.* As the latter is approximately $I/\sqrt{m}$ the forward losses may be written

$$W = aI + bI^2$$

For copper-oxide, $a = 0{\cdot}25n$ and for selenium, $0{\cdot}5n$, where n is the number of rectifying elements in series per phase. These values as indicated by Figs. 13–2 and 13–5, are of course subject to variation with temperature. The value of b is obtained from the slopes of the curves of Figs. 13–2 and 13–5. If S is the slope of the curve in volts per ampere, then $b = nS/A$, where A is the total area in square centimetres of all rectifying elements in parallel per phase. Thus

$$W = n\left(0{\cdot}25I + \frac{S_c}{A}I^2\right) \quad \text{copper-oxide}$$

$$W = n\left(0{\cdot}5I + \frac{S_s}{A}I^2\right) \quad \text{selenium}$$

it being remembered that the coefficients of both I and I^2 are lowered by an increase in temperature.

The above formulae have been developed on the assumption that the r.m.s. value of the phase current is $I/\sqrt{m}$. However, as stated on page 443, this is only justifiable if $m \geq 3$ and, for the cases where $m = 1$ or 2, I must be replaced by $\sqrt{m}I_s$, where I_s is the r.m.s. current per secondary phase. Table 13–3 gives a comparison between watts obtained by observation and calculation for a single-phase selenium rectifier ($m = 1$) composed of 30 units

TABLE 13-3
20° C.

$I_{a.c.}$	$I_{d.c.}$	$V_{a.c.}$	$W_{ob.}$	$W_{cal.}$
1·15	0·5	19·5	10·0	9·5
2·4	1·0	23·0	27·5	23·5
3·55	1·5	26·2	45·0	41·0
4·8	2·0	30·0	70·0	64·5
6·0	2·5	33·0	99·5	91·5
7·2	3·0	36·0	125	123
8·15	3·4	39·0	155	151
6·5	4·0	64·0	215	188
7·3	4·5	65·5	250	230
8·1	5·0	66·0	295	275

---

* See p. 442.

in series, each 100 cm.^{2} in area. The ambient temperature of the test was 20° C., S_s being 10 volts per amp. per cm.^{2}

It will be noted from Figs. 13–2 and 13–5 that the reverse volt/ampere characteristics differ somewhat from those of the forward direction. In the case of copper oxide the volt/ampere relationship is approximately given by

$$i = ae^2$$

and the instantaneous power loss is thus

$$\begin{aligned} ei &= ae^3 \\ &= aV_m^3 \sin^3 \theta \end{aligned}$$

where V_m is the maximum value of the reverse voltage. The mean loss over a cycle per sq. cm. of disc per phase is

$$\begin{aligned} \frac{aV_m^3}{2\pi}\int_0^\pi \sin^3 \theta d\theta &= \frac{aV_m^3}{2\pi}\int_0^\pi (\tfrac{3}{4}\sin\theta = \tfrac{1}{4}\sin 3\theta)d\theta \\ &= \frac{4\sqrt{2}a}{3\pi}V^3 \text{ watts} \end{aligned}$$

With selenium the reverse current increases more rapidly than the second power of the reverse voltage. Approximately, we have

$$i = a_1e^3$$

and the mean loss per cycle is

$$\frac{a_1V_m^3}{2\pi}\int_0^\pi \sin^4 \theta d\theta$$

$$\frac{3a_1}{4}V^4 \text{ watts}$$

where $a_1 = 10^{-7}$, approximately, when i is expressed in amps. per cm^2.

In order to obtain details of reverse losses likely to be experienced in practice, the circuit of Fig. 13–32 was set up. Each rectifier consisted of fifteen selenium units in series, each 100 cm.^{2} in area.

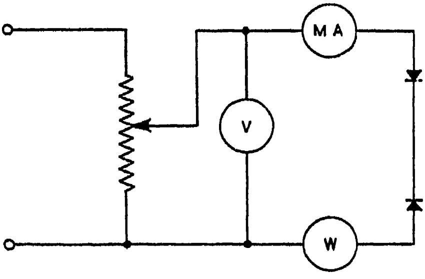


Fig. 13–32

After referring the losses to unit area, these were plotted agai maximum reverse e.m.f. per unit (r.m.s. values) and are shown Fig. 13–33. It will be noted that below about 12 volts the reve loss is practically independent of temperature and that between t figure and the normal maximum reverse voltage, 15 to 18 vo the variation with temperature is small. From Table 13–4 it evident that the reverse loss is a minimum at between 40° C. 60° C. The data from which the curves of Fig. 13–33 were c structed are given in Table 13–4. Although the reverse resista is non-linear, it is evident that the watts lost are approximat equal to the product of the reverse volts and amperes. For

TABLE 13–4

20° C.

R.M.S. Volts per Disc	R.M.S. Amp. per Cm.²	Watts per Cm.²	$\frac{\text{Volts} \times \text{Am}}{2}$ per Cm.²
8	0·065 × 10^{-3}	0·265 × 10^{-3}	0·26 × 10
10	0·11 ,, ,,	0·6 ,, ,,	0·55 ,, ,,
12	0·19 ,, ,,	1·07 ,, ,,	1·14 ,, ,,
14	0·27 ,, ,,	2·0 ,, ,,	1·9 ,,
16	0·37 ,, ,,	3·33 ,, ,,	2·96 ,, ,,
18	0·52 ,, ,,	5·0 , ,,	4·7 ,, ,,
20	0·75 ,, ,,	8·35 ,, ,,	7·5 ,, ,,
22	1·15 ,, ,,	12·0 ,, ,,	12·7 ,, ,,
24	2·2 ,, ,,	23·3 ,, ,,	26·5 ,, ,,

40° C.

R.M.S. Volts per Disc	R.M.S. Amp. per Cm.²	Watts per Cm.²	$\frac{\text{Volts} \times \text{Am}}{2}$ per Cm.²
8	0·065 × 10^{-3}	0·265 × 10^{-3}	0·26 × 10
10	0·11 ,, ,,	0·6 ,, ,,	0·55 ,, ,,
12	0·175 ,, ,,	1·0 ,, ,,	1·05 ,, ,,
14	0·245 ,, ,,	1·73 ,, ,,	1·72 ,, ,,
16	0·34 ,, ,,	2·83 ,, ,,	2·72 ,, ,,
18	0·46 ,, ,,	4·33 ,, ,,	4·12 ,, ,,
20	0·65 ,, ,,	6·85 ,, ,,	6·5 ,, ,,
22	1·0 ,, ,,	10·7 ,, ,,	11·0 ,, ,,
24	1·75 ,, ,,	18·6 ,, ,,	21·0 ,, ,,

60° C.

R.M.S. Volts per Disc	R.M.S. Amp. per Cm.^{2}	Watts per Cm.^{2}	$\frac{\text{Volts} \times \text{Amp.}}{2}$ per Cm.^{2}
8	0·065 $\times 10^{-3}$	0·265 $\times 10^{-3}$	0·26 $\times 10^{-3}$
10	0·118 ,, ,,	0·6 ,, ,,	0·59 ,, ,,
12	0·18 ,, ,,	1·07 ,, ,,	1·08 ,, ,,
14	0·255 ,, ,,	1·73 ,, ,,	1·78 ,, ,,
16	0·34 ,, ,,	2·83 ,, ,,	2·71 ,, ,,
18	0·485 ,, ,,	4·67 ,, ,,	4·36 ,, ,,
20	0·66 ,, ,,	7·0 ,, ,,	6·6 ,, ,,
22	0·1 ,, ,,	11·0 ,, ,,	11·0 ,, ,,
24	0·16 ,, ,,	17·4 ,, ,,	19·2 ,, ,,

80° C.

R.M.S. Volts per Disc	R.M.S. Amp. per Cm.^{2}	Watts per Cm.^{2}	$\frac{\text{Volts} \times \text{Amp.}}{2}$ per Cm.^{2}
8	0·09 $\times 10^{-3}$	0·33 $\times 10^{-3}$	0·36 $\times 10^{-3}$
10	0·14 ,, ,,	0·66 ,, ,,	0·7 ,, ,,
12	0·21 ,, ,,	1·26 ,, ,,	1·26 ,, ,,
14	0·295 ,, ,,	2·06 ,, ,,	2·06 ,, ,,
16	0·4 ,, ,,	3·33 ,, ,,	3·2 ,, ,,
18	0·55 ,, ,,	5·05 ,, ,,	4·95 ,, ,,
20	0·78 ,, ,,	8·0 ,, ,,	7·8 ,, ,,
22	1·15 ,, ,,	12·4 ,, ,,	12·6 ,, ,,
24	2·0 ,, ,,	22·4 ,, ,,	24·0 ,, ,,

case shown it was necessary to divide this product by 2 as two groups of discs were employed.

In order to determine the reverse losses for a given rectifier the maximum reverse voltage (r.m.s.) per unit, or disc, must first be determined. With a knowledge of the working temperature the watts per square centimetre are then found from Table 13–4. Multiplying this figure by the total area of all the discs then gives the total reverse loss. As an example we shall consider the rectifier of Fig. 13–34.

This rectifier, designed by the author, has an output of 26 volts 150 amps. at an input of 415 volts 3 phase. The circuit is a six-phase bridge, as described on page 468, connected star-star. The rectifier element consists of six units, each unit comprising two groups of twelve 112 mm. diameter discs in series-parallel, i.e. each group contains twelve discs in parallel, the two groups being in series to withstand the reverse voltage, the maximum r.m.s. value of this being 28 volts. The forward volt drop being 6 volts at full load, the forward loss is 900 watts. From Table 13–4 the watts lost per sq. cm. at 60° C. are $1{\cdot}73 \times 10^{-3}$. The total area of the discs being 14,400 sq. cms., the total reverse loss is

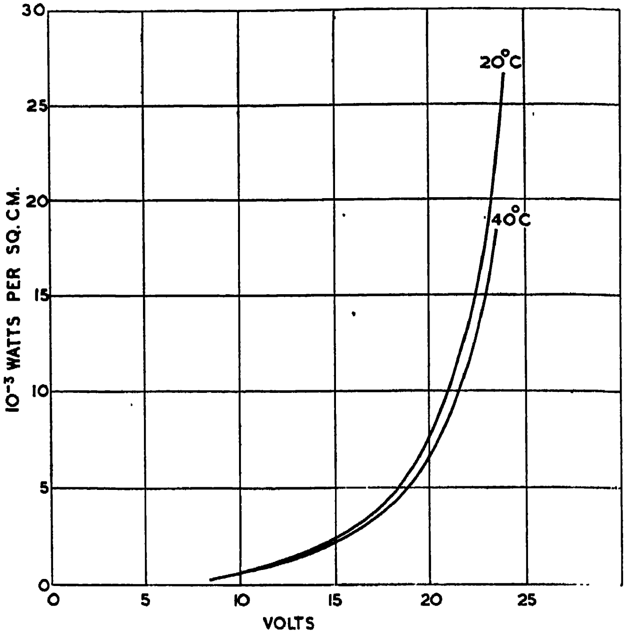


FIG. 13–33

$$14{,}400 \times 1{\cdot}73 \times 10^{-3} = 25 \text{ watts}$$

which is clearly negligible compared with the rectifier forward full-load losses.

Rectifier Voltage and Current Relationships

It is evident from foregoing details of the rectification process that a rectifier virtually constitutes a switching or commutating device, rearranging the voltage of the a.c. supply in such a manner that unidirectional impulses result. As current only flows in that phase the potential of which is higher than the remainder, then, from Fig. 13–35, it is evident that conduction commences and finishes at the intersections of the phase voltages. The resulting d.c. voltage being directly derived from the various a.c. phase voltages, we may readily formulate the relationship between them in the following manner. If V_s is the r.m.s. value of the transformer secondary phase to neutral voltage, then

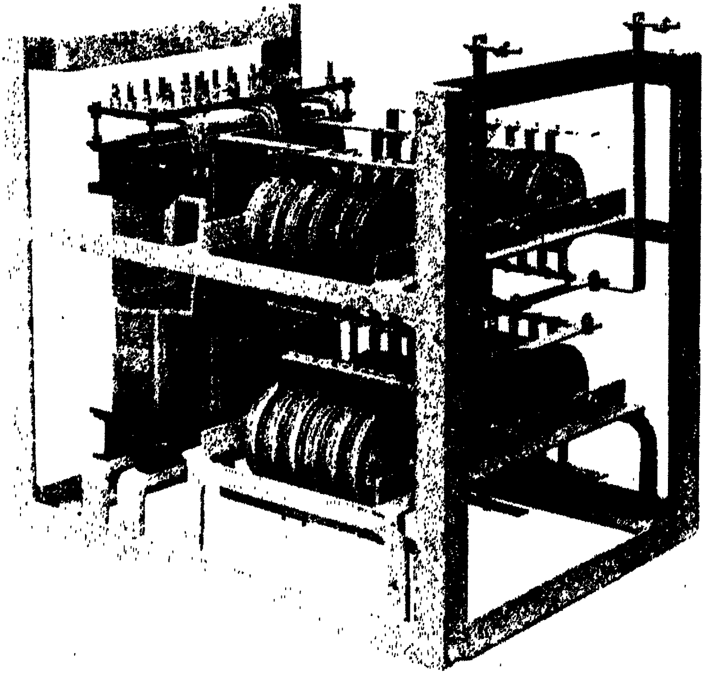


FIG. 13–34

$$e = \sqrt{2}V_s \sin \theta \quad . \quad . \quad . \quad (13\text{–}5)$$

where e is the value of the instantaneous e.m.f. The d.c. voltage, E_0, is the mean value of (13–5), and thus

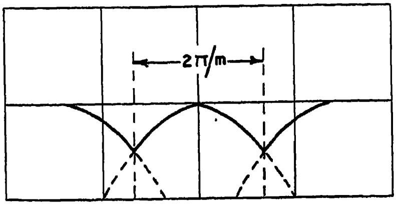


FIG. 13–35

$$E_0 = \sqrt{2}V_s \times \frac{m}{2\pi}\int_{\frac{m-2}{2m}\pi}^{\frac{m+2}{2m}\pi} \sin\theta d\theta$$

$$= \sqrt{2}V_s \cdot \frac{m}{\pi}\sin\frac{\pi}{m} \quad . \quad (13\text{–}6)$$

where m is the number of phases. Hence the requisite phase to neutral voltage for a given d.c. voltage, E_0, is

$$V_s = \frac{\pi E_0}{\sqrt{2}m \sin \pi/m} \quad . \quad . \quad . \quad . \quad (13\text{–}7)$$

If overlap be ignored, then, from Fig. 13–35, it is evident that the current persists in each phase for $2\pi/m$ of a cycle only. Thus, the mean current per phase is

$$I_a = I_m \times \frac{1}{2\pi}\int_{\frac{m-2}{2m}\pi}^{\frac{m+2}{2m}\pi} \sin\theta d\theta$$

$$= \frac{I_m}{\pi}\sin\frac{\pi}{m} \quad . \quad . \quad . \quad . \quad (13\text{–}8)$$

where I_m is the maximum value of the phase current. The total value of the d.c. current is $mI_a = I$. The r.m.s. value of the secondary phase current is

$$I_s = \sqrt{I_m^2 \times \frac{1}{2\pi}\int_{\frac{m-2}{2m}\pi}^{\frac{m+2}{2m}\pi} \sin^2\theta d\theta}$$

$$= I_m\sqrt{\frac{1}{2\pi}\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)}$$

$$= \pi I_a \cdot \frac{\sqrt{\frac{1}{2\pi}\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)}}{\sin \pi/m}$$

$$= I\pi \frac{\sqrt{\frac{1}{2\pi}\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)}}{m \sin \pi/m} \quad . \quad . \quad (13\text{–}9)$$

It will be found that for values of $m \geq 3$, (13–9) may be approximately represented by

$$I_s = I/\sqrt{m} \quad . \quad . \quad . \quad . \quad . \quad . \quad (13\text{–}10)$$

Utility Factor

The utility factor of a rectifier may be defined as the ratio of the volt-amperes obtained from the rectifier on the output side to the volt-amperes supplied on the input side. Thus, the utility factor may be taken as a measure of the degree to which the rectifier is capable of utilizing the energy supplied to it. If u.f. represents the utility factor, then

$$\text{u.f.} = \frac{E_0 I}{m V_s I_s} = \frac{1}{m} \cdot \frac{E_0}{V_s} \cdot \frac{I}{I_s}$$

$$= \frac{1}{m} \cdot \frac{\sqrt{2}m}{\pi} \sin \frac{\pi}{m} \cdot \frac{m \sin \frac{\pi}{m}}{\pi\sqrt{\frac{1}{2\pi}\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)}}$$

$$= \frac{\sqrt{2}m \sin^2 \frac{\pi}{m}}{\pi^2\sqrt{\frac{1}{2\pi}\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)}} \quad . \quad . \quad . \quad (13\text{–}11)$$

If the approximate value of the d.c. current is taken, i.e. $I_s = I/\sqrt{m}$, (13–11) becomes

$$\frac{\sqrt{2m}}{\pi} \sin \frac{\pi}{m}$$

The relation between u.f. and m is shown by Fig. 13–36. From it will be noted that the utility factor is a maximum for $m = 2{\cdot}7$. In practice, of course, only integral values of m can be taken and hence the rectifier with the largest utility factor is that which has $m = 3$.

The value given by (13–11) is sometimes referred to as the *secondary* utility factor, as the rectifier transformer primary winding has not been taken into consideration. In cases where the currents flowing in the primary are similar to those in the secondary (Figs. 14–3 and 14–6, for example) the utility factor based on a consideration of the rating of the primary circuit is the same as that based on a consideration of the secondary. However, as will appear later, certain rectifier transformer arrangements are such that the volt-ampere rating of the primary is lower than that of the secondary, with the result that the utility factor of the rectifier, referred to the primary, is higher than that given by (13–11).

The voltage relations developed on page 442 are, of course, for no-load conditions only, and are based on a sinusoidal waveform for the transformer secondary output voltage. Under load various volt drops must be considered, one of which is associated with a distortion of the transformer secondary voltage wave. A factor of considerable importance under load conditions is the transformer leakage inductance. Neglect of this quantity involves the assumption that the arc commutates from one anode to the next instantaneously. An instantaneous change of current of this character in an inductive circuit is obviously impossible. Hence, in practice, the anode current gradually decreases towards the end of the conducting period and, similarly, gradually increases when conduction commences. In many cases a smoothing choke is included in the cathode lead and a result of this is to maintain the d.c. output current free from ripple. Thus, while the current of one anode is decreasing, the current of another is increasing, the sum of the two remaining constant and equal to the load current. The co-existence of these two currents is termed *overlap*, and the number of electrical degrees over which the phenomenon persists the *angle of overlap*, u. As two anodes are simultaneously conducting, and their arc drops are the same, the voltages of the two anodes must be equal and identical with the d.c. output voltage.

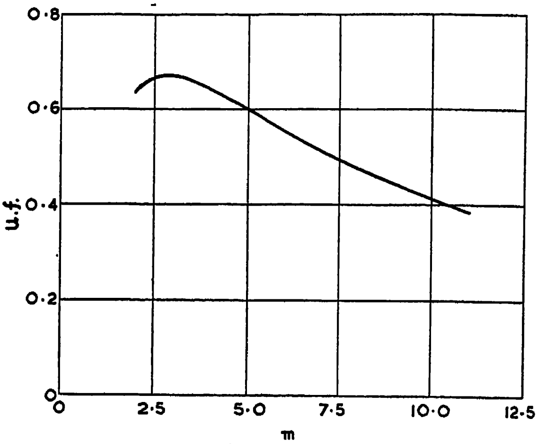


FIG. 13–36

The various conditions may be realized with the assistance of Figs. 13–37 and 13–38, which represent a hexaphase rectifier circuit. Here e_1, e_2, i_1, i_2, etc., represent the various phase voltages and currents, and r_1, r_2, x_1, x_2 the various transformer resistances and reactances. Of course, normally, we have $r_1 = r_2 = r_3$, etc., and $x = x_2 = x_3$, etc. Also $x = wL$, where L is the equivalent leakage inductance per phase of the transformer referred to the secondary.

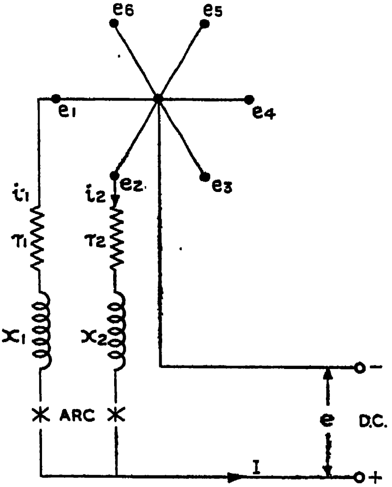


FIG. 13–37

Referring to Fig. 13–37, and assuming that resistance effects are negligible during commutation, we have, during overlap,

$$e = e_1 - L\frac{di_1}{dt} = e_2 - L\frac{di_2}{dt} \quad . \quad . \quad (13\text{–}12)$$

or

$$2e = e_1 + e_2 - L\left(\frac{di_1}{dt} + \frac{di_2}{dt}\right)$$

But

$$i_1 + i_2 = I = \text{constant} \quad . \quad . \quad . \quad (13\text{–}13)$$

where I is the d.c. load current. Hence

$$\frac{di_1}{dt} + \frac{di_2}{dt} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (13\text{–}14)$$

and

$$e = \frac{e_1 + e_2}{2}$$

Thus the output voltage, e, during the commutation period is the mean of the two anode voltages e_1, e_2.

Referring to Fig. 13–38, it will be seen that over the region *ab* anode 1 is conducting, e_1 being larger than e_2. At *b*, $e_1 = e_2$, anode 2 commences to conduct, and over the region *bc* *e* is the average of e_1 and e_2. At *c*, i_1 becomes zero, i_2 is equal to *I*, and *e* sharply rises from $\frac{1}{2}(e_1 + e_2)$ to e_2.

It is evident that an effect of overlap is to depress the rectified voltage over the period *bc* and so lower the output voltage. The degree to which this is lowered is a function of the overlap *u* and may be determined in the following manner. Referring to Fig. 13–38, and taking the intersection of e_1 and e_2 as origin, we have

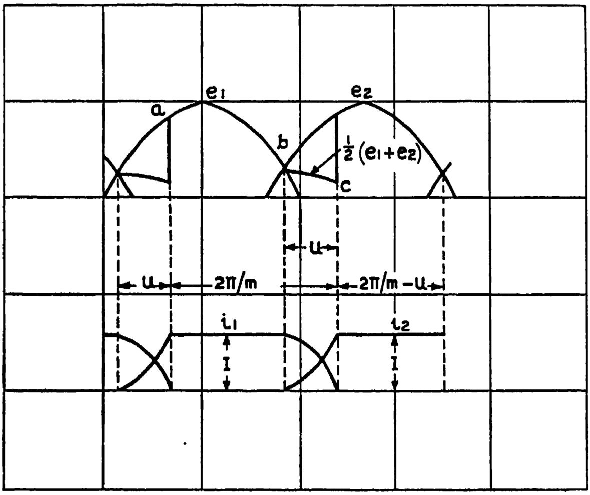


FIG. 13–38

$$E = \frac{m}{2\pi}\left[\int_0^u \frac{1}{2}(e_1 + e_2)d\theta + \int_u^{2\pi/m} e_2 d\theta\right]$$

$$= \sqrt{2}V_s \times \frac{m}{2\pi}\left[\int_0^u \frac{1}{2}\left\{\sin\left(\frac{m+2}{2m}\pi + \theta\right) + \sin\left(\frac{m-2}{2m}\pi + \theta\right)\right\}d\theta\right.$$

$$\left. + \int_u^{2\pi/m} \sin\left(\frac{m-2}{2m}\pi + \theta\right)d\theta\right] \quad . \quad (13\text{–}15)$$

$$= \sqrt{2}V_s \times \frac{m}{2\pi}\left[\sin u \cos\frac{\pi}{m} + \sin\left(\frac{\pi + u}{2}\right)\sin\left(\frac{\pi}{m} - \frac{u}{2}\right)\right]$$

$$= \sqrt{2}V_s \times \frac{m}{2\pi}\left[\sin u \cos\frac{\pi}{m} + 2\cos^2\frac{u}{2}\sin\frac{\pi}{m} - 2\cos\frac{\pi}{m}\sin\frac{u}{2}\cos\frac{u}{2}\right]$$

$$= \sqrt{2}V_s \cdot \frac{m}{\pi}\sin\frac{\pi}{m}\cos^2\frac{u}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (13\text{–}16)$$

Comparing this expression with that of (13–6), we see that when the transformer reactance is taken into consideration the rectifier voltage is given by

$$E = E_0 \cos^2\frac{u}{2} \quad . \quad . \quad . \qquad (13\text{–}17)$$

The drop in voltage due to reactance is

$$E_0\left(1 - \cos^2\frac{u}{2}\right) = E_0 \sin^2\frac{u}{2} \quad . \quad . \qquad (13\text{–}18)$$

From the foregoing it is obvious that the angle of overlap is an important factor in rectifier performance, and hence we must now determine the quantities on which its magnitude depends.

From (13–12) and (13–14) we have

$$\frac{di_1}{dt} = \frac{e_1 - e_2}{2L}$$

Also

$$e_1 - e_2 = \sqrt{2}V_s\left[\sin\left(\frac{m+2}{2m}\pi + \theta\right) - \sin\left(\frac{m-2}{2m}\pi + \theta\right)\right]$$

$$= -2\sqrt{2}V_s \sin\frac{\pi}{m}\sin\theta$$

and $\frac{di_1}{dt} = -\frac{\sqrt{2}V_s}{L}\sin\frac{\pi}{m}\sin\theta$

which leads to

$$i_1 = \frac{\sqrt{2}V_s}{wL}\sin\frac{\pi}{m}\cos\theta + C \quad (\text{as } \theta = wt)$$

where C is a constant and $w = 2\pi f$. When $t = 0$, $i_1 = I$, and hence the expression for i_1 becomes

$$i_1 = I - \frac{\sqrt{2}V_s}{X}\sin\frac{\pi}{m}(1 - \cos\theta) \quad . \quad . \quad . \qquad (13\text{–}19)$$

where $X = wL$, i.e. the transformer leakage reactance.

From (13–13)

$$i_2 = I - i_1 = \frac{\sqrt{2}V_s}{X} \sin\frac{\pi}{m}(1 - \cos\theta) \quad . \quad . \quad (13\text{–}20)$$

Now when $i_1 = 0$, $\theta = u$, and under these circumstances (13–20) becomes

$$I - \frac{\sqrt{2}V_s}{X} \sin\frac{\pi}{m}(1 - \cos u) = 0$$

from which

$$1 - \cos u = \frac{IX}{\sqrt{2}V_s \sin\frac{\pi}{m}} \quad . \quad . \quad (13\text{–}21)$$

and

$$u = \operatorname{arc}\cos\left(1 - \frac{IX}{\sqrt{2}V_s \sin\frac{\pi}{m}}\right)$$

Also

$$\cos u = 1 - 2\sin^2\frac{u}{2}$$

and, substituting in (13–21),

$$\sin\frac{u}{2} = \sqrt{\frac{IX}{2\sqrt{2}V_s \sin\pi/m}} \quad . \quad . \quad (13\text{–}22)$$

Thus, it is seen that the angle of overlap is an increasing function of I and X. Substituting from (13–22) into (13–18) we have

$$E_0 \sin^2\frac{u}{2} = \frac{E_0 IX}{2\sqrt{2}V_s \sin\pi/m}$$

which shows that the decrease in output voltage due to transformer reactance is a linear function of the load current. Hence, a rectifier possesses a straight-line regulation characteristic. Writing $\cos u = 2\cos^2\frac{u}{2} - 1$ and, substituting for $\cos^2\frac{u}{2}$ in (13–17),

$$E = E_0\left(1 - \frac{IX}{2\sqrt{2}V_s \sin\frac{\pi}{m}}\right)$$

$$= E_0 - \frac{E_0 IX}{2\sqrt{2}V_s \sin\pi/m}$$

But

$$E_0 = \sqrt{2}V_s \frac{m}{\pi} \sin\frac{\pi}{m}$$

and thus

$$E = E_0 - \frac{m}{\pi} \cdot \frac{IX}{2} \quad . \quad . \quad . \quad . \quad . \quad (13\text{–}23)$$

Hence we see that for given values of I and X the decrease in voltage under load is directly proportional to the number of phases employed.

THE EFFECTIVE VALUE OF ANODE CURRENT WITH OVERLAP

In addition to the foregoing modification of the output voltage, the overlap also modifies the r.m.s. value of the anode current as given by (13–9). Referring to Fig. 13–38 the current may be regarded as consisting of three parts: two parts, i_1 and i_2 each, extending over an angle u, and a third intermediate part, I, extending over an angle $(2\pi/m - u)$. The r.m.s. value of the current is, therefore,

$$I_s = \sqrt{\frac{1}{2\pi}\left[\int_0^u i_1^2 d\theta + \int_u^{2\pi/m} I^2 d\theta + \int_0^u i^2 d\theta\right]} \quad . \quad (13\text{–}24)$$

From (13–19) and (13–21) we may write

$$i_1 = I\left(1 - \frac{1-\cos\theta}{1-\cos u}\right)$$

also

$$i_2 = I\left(\frac{1-\cos\theta}{1-\cos u}\right)$$

Substituting in (13–24)

$$I_s = \sqrt{\frac{I^2}{2\pi}\left[\left(\frac{2\pi}{m} - u\right) + \int_0^u\left(1 - \frac{1-\cos\theta}{1-\cos u}\right)^2 d\theta + \int_0^u\left(\frac{1-\cos\theta}{1-\cos u}\right)^2 d\theta\right]}$$

$$= \sqrt{\frac{I^2}{m} + \frac{I^2}{\pi}\left[\int_0^u\left\{\left(\frac{1-\cos\theta}{1-\cos u}\right)^2 - \left(\frac{1-\cos\theta}{1-\cos u}\right)\right\} d\theta\right]}$$

$$= \sqrt{\frac{I^2}{m} + I^2\left[\frac{-(2+\cos u)\sin u + (1+2\cos u)u}{2\pi(1-\cos u)^2}\right]}$$

$$= \frac{I}{\sqrt{m}}\sqrt{1 - m\left[\frac{(2+\cos u)\sin u - (1+2\cos u)u}{2\pi(1-\cos u)^2}\right]}$$

$$= \frac{I}{\sqrt{m}}\sqrt{[1 - mf(u)]} \quad . \quad . \quad . \quad . \quad . \quad . \quad (13\text{–}25)$$

Thus, a further effect of overlap is to reduce the r.m.s. value of the anode current by a factor $\sqrt{[1 - mf(u)]}$.

It is now convenient to collect together the various equations showing the effects of transformer reactance. Thus

$$E = \sqrt{2}V_s \frac{m}{\pi} \sin \frac{\pi}{m} \left[\cos^2 \frac{u}{2}\right]$$

$$= \sqrt{2}V_s \frac{m}{\pi} \sin \frac{\pi}{m} \left[1 - \frac{IX}{2\sqrt{2}V_s \sin \pi/m}\right]$$

$$= \sqrt{2}V_s \frac{m}{\pi} \sin \frac{\pi}{m} - \left[\frac{m}{\pi}\frac{IX}{2}\right]$$

Also $$I_s = \frac{I}{\sqrt{m}} [\sqrt{1 - mf(u)}]$$

the terms in square brackets arising due to reactance. Referring to Fig. 13–39, it will be found that the following factors have been

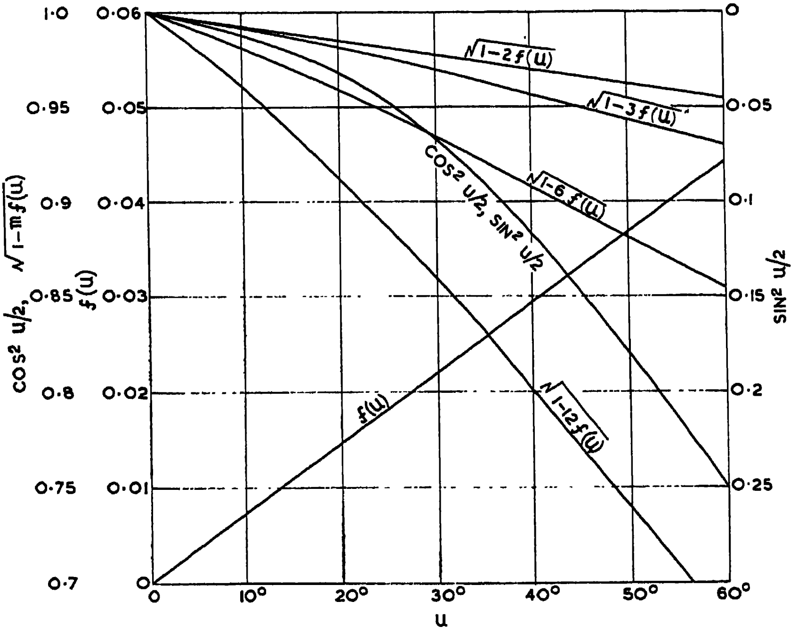


Fig. 13·39

plotted against u; i.e. $\cos^2 u/2$, $\sin^2 u/2$, and $\sqrt{1 - mf(u)}$. It will be noted that for $m > 3$, the reduction in anode current and hence transformer copper loss may be quite pronounced. However, this is seldom applicable to 12-phase rectification for (as will be shown later) in this case star-connected transformers are rarely used.

Arc and Resistance Drops

In determining the voltage drop due to reactance, the resistance of the transformer windings was neglected. This is justifiable during commutation as, due to the rapidly-changing current during this period, the resistance drop is negligible compared with the reactance drop. During the time the current is relatively constant, i.e. equal to I, the reactance drop is negligible compared with the resistance drop, and the latter must then be taken into consideration in determining the rectifier voltage. If R is the resistance per phase referred to the secondary, then the mean volt drop due to resistance is IR. Assuming the arc drop to be constant and equal to E_a, the total rectifier drop is

$$\frac{m}{\pi}\cdot\frac{IX}{2}+IR+E_a$$

and, finally,

$$E=\sqrt{2}V_s\frac{m}{\pi}\sin\frac{\pi}{m}-\left(\frac{m}{\pi}\cdot\frac{IX}{2}+IR+E_a\right) \quad . \quad (13\text{–}26)$$

Under no-load conditions

$$E_{nl}=2V_s\frac{m}{\pi}\sin\frac{\pi}{m}-E_a$$

and the percentage regulation is given by

$$100\times\left(\frac{E_{nl}-E}{E_{nl}}\right)$$

$$=100\times\frac{\dfrac{m}{\pi}\cdot\dfrac{IX}{2}+IR}{E_0-E_a}$$

It will be appreciated from (13–26) that when, in addition to the reactance drop, the resistance and arc drops are considered, a rectifier still possesses a straight-line regulation characteristic.

BIBLIOGRAPHY

"Electrolytic Rectifiers," F. G. Spreadbury, *The British Engineer*, April, 1938, p. 53.

"Low Power Rectification," F. G. Spreadbury, *The Electric Power Engineer*, Nov., 1938, p. 782.

CHAPTER XIV

RECTIFIER CIRCUITS

THERE are possibly in existence some forty to fifty different rectifier circuits for connecting rectifiers to alternating current supplies. These circuits are, in general, either single- or three-phase on the primary side, the majority of circuits appertaining to the latter. The circuits are of varying merit, and, in practice, it is probable that not more than a dozen different types are used. In the following analysis only the more commonly employed circuits will be considered, the undermentioned assumptions initially being made to clarify the treatment—

1. The alternating supply voltage is taken to be sinusoidal.
2. The various volt drops are ignored.
3. The magnetizing current of the rectifier transformer is assumed to be negligible.

Commenting on these assumptions, with modern supplies, the first is perfectly reasonable; with reference to (2), allowance is easily made for the various drops from previous analyses; the effect of magnetizing current is principally to lower the power factor of the rectifier, particularly at light loads. It may be shown, however, that the influence of the transformer magnetizing current on the rectifier power-factor can be readily calculated.

Single-phase, Half-wave Circuit

The simplest form of rectifier circuit is shown by Fig. 14–1, this giving what is known as half-wave rectification. Actually we have a one-phase system and, therefore, $m = 1$. However, the conducting period persists as for $m = 2$, and thus this value must be employed in the sine term of (13–7). As there is only one phase, however, the factor m in (13–7) must be taken as 1. Hence the secondary phase voltage is given by

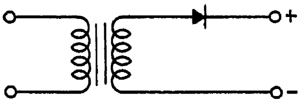


FIG. 14–1

$$V_s = \frac{\pi E_0}{\sqrt{2}\sin \pi/2} = 2{\cdot}22E_0 \quad . \quad . \quad (14\text{–}1)$$

The same considerations apply in determining the secondary phase current, and we have

$$I_s = \frac{I\pi\sqrt{\frac{1}{2\pi}\left(\frac{\pi}{2} + \frac{1}{2}\sin\pi\right)}}{\sin\pi/2} = 1{\cdot}57I \qquad . \qquad (14\text{–}2)$$

The peak value of the phase current is

$$I_m = \frac{\pi}{m\sin\pi/m}I = 3{\cdot}14I$$

Now when current flows in the secondary phase, the ampere-turns due to this must be balanced by an equal number on the primary side, and hence the primary current I_p is given by $(N_s/N_p)I_s$, where N_s and N_p are, respectively, the numbers of turns on the secondary and primary windings. It is evident that the line current is equal in value to the primary phase current.

Transformer Secondary Rating

From (14–1) and (14–2) it is apparent that the volt-ampere rating of the secondary winding is

$$V_sI_s = 2{\cdot}22E_0 \times 1{\cdot}57I = 3{\cdot}5E_0I \qquad .$$

the primary winding also having the same rating.

A.C. Supply Loading

The primary line voltage V_L is the same as the primary phase voltage V_p. Also the line and primary phase currents are identical. It follows that the line volt-ampere input is

$$\begin{aligned} V_LI_L &= V_pI_p = (N_p/N_s)V_s \times (N_s/N_p)I_s \\ &= V_sI_s = 3{\cdot}5E_0I \end{aligned}$$

Voltage Fluctuation

The d.c. output voltage periodically fluctuates between $\sqrt{2}V_s$ and zero. Hence the fluctuation is $\sqrt{2}V_s$ or $3{\cdot}12E_0$.

Single-phase, Full-wave Circuit

Two different forms of full-wave circuits exist, the first of which is shown by Fig. 14–2. In this case $m = 2$ and

$$V_s = \frac{\pi E_0}{2\sqrt{2}\sin\pi/2} = 1{\cdot}11E_0 \qquad . \quad . \quad . \qquad (14\text{–}3)$$

Also
$$I_s = \frac{\pi\sqrt{\dfrac{1}{2\pi}\left(\dfrac{\pi}{2} + \tfrac{1}{2}\sin\pi\right)}}{2\sin\pi/2}\,I = 0{\cdot}785I \quad . \quad (14\text{–}4)$$

The peak value of the secondary phase current is

$$I_m = \frac{\pi}{2\sin\pi/2}\,I = 1{\cdot}57I \quad . \quad . \quad (14\text{–}5)$$

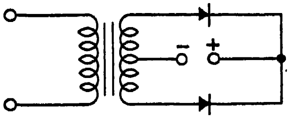


FIG. 14–2

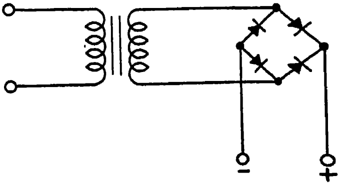


FIG. 14–3

With this circuit it will be noted that the transformer primary winding carries current twice per cycle as two secondary phases are successively employed. Hence the primary current is

$$I_p = \sqrt{2}(N_s/N_p)I_s = 1{\cdot}11(N_s/N_p)I. \quad . \quad (14\text{–}6)$$

Transformer Secondary Rating

From (14–3) and (14–4) the secondary rating is

$$2V_sI_s = 2 \times 1{\cdot}11E_0 \times 0{\cdot}785I = 1{\cdot}74E_0I$$

Transformer Primary Rating

From (14–6) the primary rating is

$$\begin{aligned} V_pI_p &= (N_p/N_s)V_s \times 1{\cdot}11(N_s/N_p)I \\ &= 1{\cdot}11 \times 1{\cdot}11E_0I = 1{\cdot}23E_0I \end{aligned}$$

Thus the mean volt-ampere rating of the transformer is

$$(1{\cdot}74 + 1{\cdot}23)E_0I/2 = 1{\cdot}48E_0I$$

A.C. Supply Loading

In this case it is evident that the a.c. supply loading is equal to the transformer primary rating, i.e. $1{\cdot}23E_0I$.

Power Factor

On the assumption that a sufficiently large cathode choke is employed to maintain the output current constant, the power output is constant and equal to E_0I. Ignoring losses, this is also the input and hence the power factor is given by the ratio of the output to the volt-ampere input; i.e.

$$\frac{E_0I}{1{\cdot}23E_0I} = 0{\cdot}81$$

Voltage Fluctuation

As with the previous circuit, the voltage fluctuates between $\sqrt{2}V_s$ and zero. Thus the magnitude of the fluctuation is $1{\cdot}11\sqrt{2}E_0 = 1{\cdot}57E_0$.

Maximum Inverse Voltage

Ignoring the drop in the conducting anode, the non-conducting anode must withstand twice the phase to neutral voltage, the maximum value of which is $2\sqrt{2}V_s$.

The Bridge Circuit

The second form of single-phase full-wave rectifier circuit referred to above is shown by Fig. 14–3 and is known as the bridge or Graetz circuit. Contrasting this with the circuit of Fig. 14–2, it will be noted that the bridge circuit permits full-wave rectification to be obtained from a two-terminal source, a tapped secondary winding being unnecessary. The method by which rectification is effected is self-evident. A disadvantage of the bridge circuit when low voltages are to be rectified is that four rectifier elements are needed instead of two. Further, it is apparent that the direct current must traverse two units in series against one for the circuit of Fig. 14–2. This, of course, makes for increased regulation and losses. However, there are several advantages to be set in favour of the bridge circuit. For a given d.c. output voltage the inverse peak voltage per rectifier in the circuit of Fig. 14–2 is twice that of Fig. 14–3. Hence if the inverse voltage is considerably in excess of that which can be withstood by a single-rectifier element, the number of elements in the circuits of Figs. 14–2 and 14–3 will be the same. This, of course, is generally the case with metal rectifiers where the maximum reverse voltage per disc is relatively low. In the circuit analysis which follows it will be found that the particular advantage of the bridge circuit is the reduced rating of the transformer as compared with that of Fig. 14–2.

The circuit of Fig. 14–3 is regarded as a two-phase rectifier, and hence $m = 2$. Then

$$V_s = \frac{\pi E_0}{2\sqrt{2}\sin \pi/2} = 1{\cdot}11E_0 \quad . \quad . \quad (14\text{–}7)$$

However, the single secondary winding supplies current twice per cycle, and thus the r.m.s. value of the secondary current is $\sqrt{2}$ times that given by (14–4), i.e. $1{\cdot}11I$.

The peak value of the secondary phase current is

$$I_m = \frac{\pi}{2\sin \pi/2} I = 1{\cdot}57I$$

As the primary and secondary windings have the same number of current pulses per cycle,

$$I_p = (N_s/N_p)I_s = 1{\cdot}11(N_s/N_p)I$$

Transformer Secondary Rating

The secondary rating is

$$V_sI_s = 1{\cdot}11E_0 \times 1{\cdot}11I = 1{\cdot}23E_0I$$

the primary rating having also this value.

A.C. Supply Loading

As with Fig. 14–2, the supply loading is equal to the primary rating, i.e. $1{\cdot}23E_0I$.

Power Factor

The power factor is given by $E_0I/1{\cdot}23E_0I = 0{\cdot}81$.

Voltage Fluctuation

This is the same as for the circuit of Fig. 14–2, i.e. $1{\cdot}57E_0$.

Maximum Inverse Voltage

In this case the peak inverse voltage is one-half of that for the circuit of Fig. 14–2, i.e. $\sqrt{2}V_s$.

Three-phase Rectification

A disadvantage of single-phase rectification is, of course, the large voltage fluctuation or ripple which exists in the output voltage. By employing a three-phase supply and rectification this fluctuation may be considerably reduced as indicated by Fig. 14–4. Moreover, by the employment of special windings and connexions on the transformer secondary side, six- or twelve-phase (or even more) rectification may be produced with a corresponding reduction in ripple. However, with a three-phase supply particular attention must be paid to the primary connexions (whether star or delta), as the following example will show.

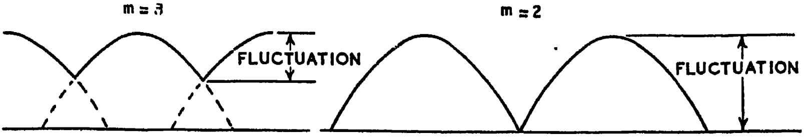


FIG. 14–4

Considering the star-star arrangement of Fig. 14–5, when one anode is conducting, a secondary phase, such as *ab*, is loaded while the other two are unloaded. Due to mutual inductance between *ab* and *AB*, the impedance of *AB* falls, while that of *AD* and *AC* remains high. But the algebraical sum of the currents at *A* must equal zero and therefore the sum of the currents in *AD* and *AC* must equal that in *AB*. Hence the voltages on *AD* and *AC* must be much higher than on *AB*, with the result that the neutral moves along *AB*. This phenomenon is evidently productive of very poor voltage regulation and, in addition, leads to distortion of the primary supply current.

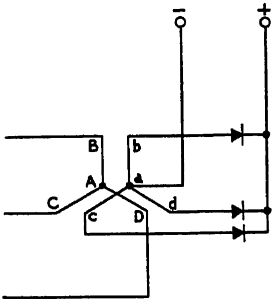


FIG. 14–5

The slipping of the neutral, as it is termed, may be prevented by fitting an extra mesh-connected tertiary winding to the transformer. This tends to fix the neutral point by reason of the fact that any unbalancing due to a load in *ab* will cause a current to circulate in the tertiary winding, which will affect all primary phases, tending to keep their impedances equal. However, a tertiary winding is an additional cost and may be avoided by a mesh-connected primary or certain secondary connexions to be described later.

THREE-PHASE DELTA-STAR CIRCUIT

This circuit is shown by Fig. 14–6. As $m = 3$ we have

$$V_s = \frac{\pi E_0}{3\sqrt{2}\sin \pi/3} = 0{\cdot}86E_0$$

The secondary phase current is

$$I_s = \frac{\pi\sqrt{\dfrac{1}{2\pi}\left(\dfrac{\pi}{3} + \dfrac{1}{2}\sin\dfrac{2\pi}{3}\right)}}{3\sin\pi/3}\,I = 0{\cdot}58I$$

The peak value of which is

$$I_m = \frac{\pi}{3\sin\pi/3}\,I = 1{\cdot}21I$$

The primary phase current is $(N_s/N_p)I_s = 0{\cdot}58(N_s/N_p)I$.

Transformer Secondary Rating

The secondary rating is

$$3V_sI_s = 3 \times 0{\cdot}86E_0 \times 0{\cdot}58I = 1{\cdot}5E_0I$$

the primary rating having the same value.

A.C. Supply Loading

Referring to Fig. 14–6, it will be appreciated that each line must carry the phase current twice per cycle, once for each phase connected to the line. Hence the line current is

$$\sqrt{2}(N_s/N_p)I_s = 0{\cdot}82(N_s/N_p)I$$

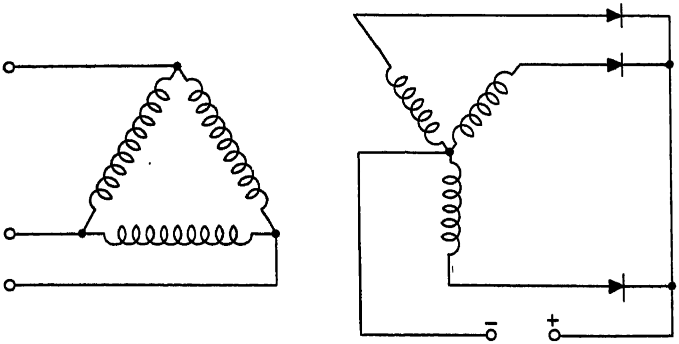


FIG. 14–6

The line voltage is the same as the phase voltage and thus the line loading is

$$\begin{aligned}\sqrt{3}V_LI_L &= \sqrt{3}(N_p/N_s)V_s \times 0{\cdot}82(N_s/N_p)I\\ &= \sqrt{3} \times 0{\cdot}86E_0 \times 0{\cdot}82I\\ &= 1{\cdot}22E_0I\end{aligned}$$

Power Factor

The power factor is given by

$$\frac{E_0 I}{1{\cdot}22 E_0 I} = 0{\cdot}82$$

Voltage Fluctuation

In this case the output voltage does not fall to zero as with single-phase circuits, the fluctuation occurring between $\sqrt{2}V_s$ and $\sqrt{2}V_s \sin \pi/6 = V_s/\sqrt{2}$. The magnitude of the fluctuation is thus

$$\sqrt{2}V_s - V_s/\sqrt{2}$$
$$= \left(\sqrt{2} - \frac{1}{\sqrt{2}}\right)V_s = 0{\cdot}61E_0$$

Maximum Inverse Voltage

Ignoring the drop in the conducting anode, the non-conducting anodes must withstand $\sqrt{3}$ times the phase to neutral voltage, i.e. $\sqrt{6}V_s$.

THREE-PHASE INTER-STAR CIRCUIT

The three-phase circuit to be now described is shown by Fig. 14–7 and makes use of somewhat special connexions on the secondary side. The secondary winding consists of six identical coils connected in the manner shown to form a three-phase supply. As any pair of coils forming a phase are located on different limbs of the transformer, it follows that two primary phases are simultaneously carrying currents of the same amount. Hence the algebraical sum of the primary currents is equal to zero, and either a delta- or star-connected primary winding may be employed. Further, each primary phase is utilized twice per cycle, which results in a reduction in the primary rating as compared with that of Fig. 14–6. This reduction, however, is partially

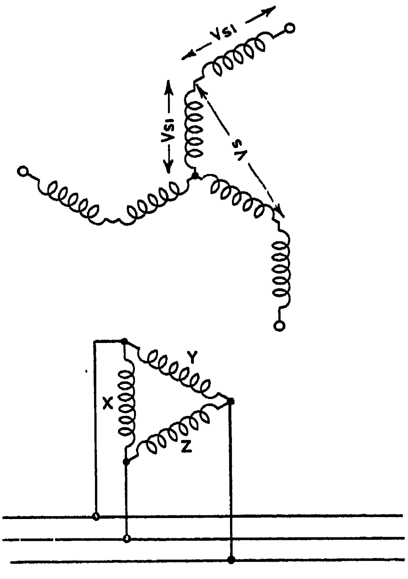


FIG. 14–7

offset by an increase in the secondary rating (due to the inter-star connexion) for the phase voltage is only $\sqrt{3}/2$ of what it would be if the pair of connected windings were located on the same limbs instead of on two different limbs.

As with the circuit of Fig. 14–6, $m = 3$, and we have

$$V_s = 0{\cdot}86E_0$$
$$I_s = 0{\cdot}58I$$
$$I_m = 1{\cdot}21I$$

Now, if N_{s1} and V_{s1} are respectively the number of turns and voltage of each of the six secondary windings, then

$$V_s = \sqrt{3}V_{s1}$$

and
$$V_{s1} = V_s/\sqrt{3} = 0{\cdot}5E_0$$

As previously stated, each primary phase carries current twice per cycle and thus the primary current is given by

$$I_p = \sqrt{2}(N_{s1}/N_p)I_s = 0{\cdot}82(N_{s1}/N_p)I$$

Transformer Secondary Rating

The secondary rating is

$$6V_{s1}I_s = 6 \times 0{\cdot}5E_0 \times 0{\cdot}58I = 1{\cdot}74E_0I$$

Transformer Primary Rating

The primary rating is

$$\begin{aligned} 3V_pI_p &= 3(N_p/N_{s1})V_{s1} \times 0{\cdot}82(N_{s1}/N_p)I \\ &= 3 \times 0{\cdot}5E_o \times 0{\cdot}82I = 1{\cdot}23E_0I \end{aligned}$$

The mean rating of the transformer is thus $1{\cdot}48E_0I$.

A.C. Supply Loading

Assuming a star-connected primary is employed, the line current is equal to the phase current, while the line voltage is $\sqrt{3}$ times the phase voltage. Thus the supply loading is

$$\sqrt{3}V_LI_L = 3V_pI_p = 1{\cdot}23E_0I$$

If a delta-connected winding is employed, a line will carry the current for two primary phases (as two are always simultaneously in operation) for one-third of a cycle and then the current for two phases in succession, each for one-third of a cycle. The instantaneous value of the latter current is, of course, one-half of that of the former. Let i_p be the instantaneous value of the primary current. Then the line current is given by

$$\sqrt{\frac{1}{2\pi}\left[\int_{\pi/6}^{5\pi/6}(2i_p)^2 d\theta + \int_{\pi/6}^{5\pi/6} i^2{}_p d\theta + \int_{\pi/6}^{5\pi/6} i_p{}^2 d\theta\right]}$$

$$= \sqrt{\frac{6}{2\pi}\int_{\pi/6}^{5\pi/6} i_p{}^2 d\theta} \quad \text{or} \quad \sqrt{\frac{3\times 2}{2\pi}\int_{\pi/6}^{5\pi/6} i_p{}^2 d\theta}$$

But $$I_p = \sqrt{\frac{2}{2\pi}\int_{\pi/6}^{5\pi/6} i_p{}^2 d\theta}$$

from which $$I_L = \sqrt{3}I_p = 1{\cdot}42(N_{s1}/N_p)I$$

The supply loading is, of course, the same as with a star-connected primary.

For the foregoing circuit the power factor, voltage fluctuation, and maximum inverse voltage are identical with those of the simple three-phase circuit of Fig. 14–6.

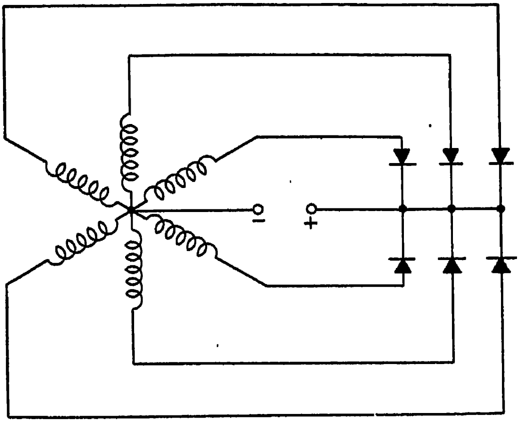


FIG. 14–8

Six-phase Rectification

The disadvantages associated with the foregoing methods of rectification are fluctuation in the d.c. output voltage and poor power factor on the input side. As previously stated, when a three-phase supply is available, it is possible to employ six- or twelve-phase connexions on the secondary side. In this way, d.c. voltage fluctuation may be considerably reduced and the power factor improved over that obtaining with three-phase rectification.

THE DELTA-DOUBLE STAR CIRCUIT

The simplest form of six-phase rectifier connexion is shown by Fig. 14–8. Each transformer limb carries two secondary windings connected in the manner shown. As only one secondary phase is in action at a time, a delta-connected primary must be employed. As $m = 6$, we have

$$V_s = \frac{\pi E_0}{6\sqrt{2}\sin\pi/6} = 0{\cdot}74E_0$$

$$I_s = \frac{\pi\sqrt{\dfrac{1}{2\pi}\left(\dfrac{\pi}{6}+\dfrac{1}{2}\sin\dfrac{2\pi}{6}\right)}}{6\sin\pi/6}\,I = 0{\cdot}41I$$

$$I_m = \frac{\pi}{6\sin\pi/6}\,I = 1{\cdot}05I$$

As each primary phase supplies two secondary phases in succession, the primary current is

$$I_p = \sqrt{2}(N_s/N_p)I_s = 0{\cdot}58(N_s/N_p)I$$

Transformer Secondary Rating

The secondary rating is

$$6V_sI_s = 6 \times 0{\cdot}74E_0 \times 0{\cdot}41I = 1{\cdot}84E_0I$$

Transformer Primary Rating

The transformer primary rating is

$$\begin{aligned}3V_pI_p &= 3(N_p/N_s)V_s \times 0{\cdot}58(N_s/N_p)I\\ &= 3 \times 0{\cdot}74E_0 \times 0{\cdot}58I = 1{\cdot}29E_0I\end{aligned}$$

The mean rating of the transformer is thus $1{\cdot}57E_0I$.

A.C. Supply Loading

As each primary phase has two equal current pulses per cycle, each line will have four equal current pulses per cycle, each of which is equal in magnitude to that of the primary current. Hence the line current is

$$I_L = \sqrt{2}I_p = 0{\cdot}82(N_s/N_p)I$$

The supply loading is, therefore,

$$\begin{aligned}\sqrt{3}V_LI_L &= \sqrt{3}(N_p/N_s)V_s \times 0{\cdot}82(N_s/N_p)I\\ &= \sqrt{3} \times 0{\cdot}74E_0 \times 0{\cdot}82I\\ &= 1{\cdot}05E_0I\end{aligned}$$

A disadvantage of the foregoing circuit is the high rating of the secondary winding of the transformer. Even when the mean rating is considered, it is seen that it is $1{\cdot}57E_0I$ as against $1{\cdot}5E_0I$ for the circuit of Fig. 14–6. However, when the reduction in ripple is considered and the fact that the power factor of Fig. 14–8 is 0·95 as against 0·81 for Fig. 14–6, the advantages of six-phase working are sufficiently evident.

Power Factor

The power factor is

$$\frac{E_0I}{1{\cdot}05E_0I} = 0{\cdot}95$$

Voltage Fluctuation

In this case the output voltage fluctuates between $\sqrt{2}V_s$ and $\sqrt{2}V_s \sin \pi/3$. The amplitude of the fluctuation is

$$\begin{aligned}&\sqrt{2}V_s - \sqrt{2}V_s \sin \pi/3\\ &= \left(\sqrt{2} - \sqrt{\frac{3}{2}}\right) V_s = 0{\cdot}14E_0\end{aligned}$$

Maximum Inverse Voltage

As with a two-phase circuit, a non-conducting anode must withstand a peak voltage $2\sqrt{2}V_s$ volts.

We shall now consider the influence of the transformer reactance on the various currents and voltages, etc., of the foregoing circuit. From (13–16)

$$V_s = 0{\cdot}74E\left[\frac{1}{\cos^2\frac{u}{2}}\right]$$

while, from (13–25), $I_s = 0{\cdot}41I\sqrt{[1 - 6f(u)]}$

Also $I_p = 0{\cdot}58(N_s/N_p)\sqrt{I[1 - 6f(u)]}$

The secondary rating is $1{\cdot}84EI\left[\frac{\sqrt{1 - 6f(u)}}{\cos^2\frac{u}{2}}\right]$

and the primary rating $1{\cdot}29EI\left[\dfrac{\sqrt{1-6f(u)}}{\cos^2\dfrac{u}{2}}\right]$

The line current is $0{\cdot}82(N_s/N_p)I[\sqrt{1-3f(u)}]$

and the supply loading $1{\cdot}05EI\left[\dfrac{\sqrt{1-3f(u)}}{\cos^2\dfrac{u}{2}}\right]$

The power factor is $0{\cdot}95\left[\dfrac{\cos^2\dfrac{u}{2}}{\sqrt{1-3f(u)}}\right]$

From (13–23) $E = E_0 - \left[\dfrac{3}{\pi}IX\right]$

Expressing I in terms of I_p we may write

$$E = E_0 - \left[\frac{1{\cdot}64I_pX}{(N_s/N_p)\sqrt{1-6f(u)}}\right]$$

For the ultimate purpose of comparing the regulation of this circuit with others, we shall now assume a 1 : 1 transformation ratio. In these circumstances the reactance referred to the primary is identical with that referred to the secondary and X may be replaced by X_p. Hence the last expression becomes

$$\begin{aligned} E &= \left(E_0 - \left[\frac{1{\cdot}64I_pX_p}{\sqrt{1-6f(u)}}\right]\right) \\ &= 1{\cdot}35\left(V_p - \left[\frac{1{\cdot}22I_pX_p}{\sqrt{1-6f(u)}}\right]\right) \quad . \quad . \quad (14\text{–}8) \end{aligned}$$

and the percentage regulation is $1{\cdot}22/\sqrt{1-6f(u)}$ times the percentage transformer reactance.

THE SIX-PHASE FORK OR TRIPLE-STAR CIRCUIT

The following circuit, shown by Fig. 14–9, is widely employed, for, compared with Fig. 14–8, it has a lower volt-ampere rating and permits the use of a star-connected primary winding. As before, $m = 6$ and

$$V_s = 0{\cdot}74E_0$$
$$I_a = 0{\cdot}41I$$
$$I_m = 1{\cdot}05I$$

Each of the nine secondary windings has the same voltage rating, i.e. approximately the same number of turns, the six various phase voltages being produced by two similar windings located on different limbs of the transformer. If N_{s1} and V_{s1} are, respectively, the number of turns and voltage of each of the nine secondary windings, then

$$V_s = \sqrt{3}V_{s1}$$

and $V_{s1} = V_s/\sqrt{3} = 0{\cdot}426E_0$

The anode current and that in the outer branches of the secondary winding are found from (14–2) and are

$$I_a = 0{\cdot}41I$$

The inner branch or centre star windings carry the anode current twice in succession per cycle and hence the value of the current in these windings is

$$I_s = \sqrt{2} \times 0{\cdot}41I = 0{\cdot}58I$$

Referring to Fig. 14–9, each primary phase must carry four equal pulses of current per cycle, two for a winding such as 6–1 of the centre star and two for outer branch windings such as 3 and 4. Thus the primary current is

$$I_p = \sqrt{2}(N_{s1}/N_p)I_s$$
$$= 0{\cdot}82(N_{s1}/N_p)I$$

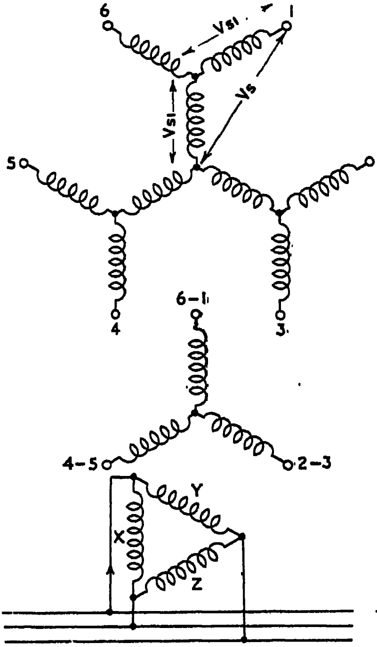


FIG. 14–9

Transformer Secondary Rating

The transformer secondary rating is

$$6V_{s1}I_a + 3V_{s1}I_s$$
$$= 6 \times 0{\cdot}426E_0 \times 0{\cdot}41I + 3 \times 0{\cdot}426E_0 \times 0{\cdot}58I$$
$$= 1{\cdot}05E_0I + 0{\cdot}74E_0I$$
$$= 1{\cdot}79E_0I$$

Transformer Primary Rating

The primary rating is

$$\begin{aligned} 3V_pI_p &= 3(N_p/N_{s1})V_{s1} \times 0{\cdot}82(N_{s1}/N_p)I \\ &= 3 \times 0{\cdot}426E_0 \times 0{\cdot}82I \\ &= 1{\cdot}05E_0I \end{aligned}$$

The mean rating is thus $1{\cdot}42E_0I$.

A.C. Supply Loading

In order to determine the line current, we may note that, when anode 1 is conducting, the line connected to the top of phase X carries the primary phase current for one-sixth of a cycle. When the arc shifts to anode 2, the line then carries the current of phase Y. When anode 3 conducts, the line then carries twice the preceding amount of current, i.e. the currents for phases X and Y. Continuing in this manner for a cycle, the line current is given by

$$\begin{aligned} I_L &= \sqrt{\frac{1}{2\pi}\left[\int_{\pi/3}^{2\pi/3} i_p{}^2d\theta + \int_{\pi/3}^{2\pi/3} i_p{}^2d\theta + \int_{\pi/3}^{2\pi/3} (2i_p{}^2)d\theta + \int_{\pi/3}^{2\pi/3} i_p{}^2d\theta \right.} \\ &\qquad\qquad \left. + \int_{\pi/3}^{2\pi/3} i_p{}^2d\theta + \int_{\pi/3}^{2\pi/3} (2i_p{}^2)d\theta \right] \\ &= \sqrt{\frac{12}{2\pi}\int_{\pi/3}^{2\pi/3} i_p{}^2d\theta} = \sqrt{\frac{3 \times 4}{2\pi}\int_{\pi/3}^{2\pi/3} i_p{}^2d\theta} \end{aligned}$$

But
$$I_p = \sqrt{\frac{4}{2\pi}\int_{\pi/3}^{2\pi/3} i_v{}^2d\theta}$$

from which
$$I_L = \sqrt{3}I_p = 1{\cdot}42(N_{s1}/N_p)I$$

The supply loading is

$$\begin{aligned} \sqrt{3}V_LI_L &= \sqrt{3}(N_p/N_{s1})V_{s1} \times 1{\cdot}42(N_{s1}/N_p)I \\ &= 3 \times 0{\cdot}426E_0 \times 1{\cdot}42I \\ &= 1{\cdot}05E_0I \end{aligned}$$

The line current derived above, of course, refers to a delta-connected primary. If a star-connected primary is employed, the line current is equal to the phase current, the supply loading being the same in both cases.

For the foregoing circuit the power factor, voltage fluctuation, and maximum inverse voltage are the same as those of the simple six-phase circuit of Fig. 14–8.

In considering the effect of transformer reactance on the various currents, we have for the anode current

$$I_a = 0{\cdot}41I\sqrt{1 - 6f(u)}$$

For the current in the centre-star winding we must take $m = 3$, as this winding carries the current of two consecutive anodes. Thus, in this case,

$$I_s = 0{\cdot}58I\sqrt{1 - 3f(u)}$$

The value of u tends to be small for the triple-star circuit, as only one secondary winding is commutated at a time. Thus, when the current changes from anode 6 to anode 1, the current in the corresponding winding of the central star does not change. The effective reactance is, therefore, the same as with a double-star circuit, the secondary voltage of which is equal to $V_s/\sqrt{3}$. When the current changes from anode 1 to anode 2 the sum of the currents in windings 1 and 2 remains constant. Hence, as these windings are located on the same limb of the transformer, the net reactance e.m.f. in each winding is zero. Hence the reactance e.m.f. of the secondary is again equal to that of a transformer, the phase to neutral voltage of which is $V_s/\sqrt{3}$. As the reactance varies as the square of the number of turns, and hence as the square of the voltage rating, the reactance to be employed in the determination of u is $X/3$, where X is the effective leakage reactance measured between the secondary transformer neutral and one anode terminal.

From (13–23) $\quad E = E_0 - \left[\frac{6}{\pi}\cdot\frac{IX}{2}\right]$

Remembering that X must be replaced by $X/3$, we have

$$E = E_0 - \left[\frac{IX}{\pi}\right]$$

Expressing I in terms of I_p

$$I_p = 0{\cdot}82(N_{s1}/N_p)I[\sqrt{1 - 3f(u)}]$$

and $\quad I = \frac{1{\cdot}22I_p}{(N_{s1}/N_p)}\left[\frac{1}{\sqrt{1 - 3f(u)}}\right]$

Assuming a 1 : 1 transformation ratio, $N_{s1} = N_p/\sqrt{3}$ and

$$I = 2{\cdot}1 I_p \left[\frac{1}{\sqrt{1 - 3f(u)}}\right]$$

Thus

$$E = \left(E_0 - \frac{2{\cdot}1 I_p X_p}{\pi\sqrt{1 - 3f(u)}}\right)$$

$$= 1{\cdot}35\left(V_p - \left[\frac{I_p X_p}{2\sqrt{1 - 3f(u)}}\right]\right) \quad . \quad (14\text{–}9)$$

The secondary rating is

$$6 \times 0{\cdot}426 \frac{E}{\cos^2\frac{u}{2}} \times 0{\cdot}41 I\sqrt{1 - 6f(u)} + 3 \times 0{\cdot}426 \frac{E}{\cos^2\frac{u}{2}} \times 0{\cdot}58 I\sqrt{1 - 3f(u)}$$

$$= 1{\cdot}05 EI \frac{\sqrt{1 - 6f(u)}}{\cos^2\frac{u}{2}} + 0{\cdot}74 EI \frac{\sqrt{1 - 3f(u)}}{\cos^2\frac{u}{2}}$$

$$= 1{\cdot}05 EI \left[\frac{\sqrt{2}\sqrt{1 - 6f(u)} + \sqrt{1 - 3f(u)}}{\sqrt{2}\cos^2\frac{u}{2}}\right]$$

and the primary rating

$$1{\cdot}05 EI \left[\frac{\sqrt{1 - 3f(u)}}{\cos^2\frac{u}{2}}\right]$$

the supply loading having the same value as the latter.

Comparing the triple-star circuit with that of the double-star, the mean transformer rating of the former is about 9 per cent lower than the latter. Furthermore, by comparing (14–9) with (14–8), it will be found that the regulation of the triple-star circuit is only about 41 per cent of that of the double-star.

THE SIX-PHASE BRIDGE CIRCUIT

Where rectifiers are constructed from a number of separate elements (such as is the case with metal and valve rectifiers), the circuit of Fig. 14–10 is frequently employed. From this it will be noted that two groups of rectifiers are star-connected, the star point of one group being the junction of its cathodes, while the junction of the anodes of the other group forms a second star point. With this circuit it will be found that six-phase rectification is obtained from a three-terminal source. In order to follow the operation of the circuit, a somewhat different method of approach will be taken compared with methods employed with foregoing circuits.

Considering the moment when the voltage of phase C is zero, the voltages of phases A and B are, respectively, $+\sqrt{3}/2$ and $-\sqrt{3}/2$ of their maximum values, i.e. they are both acting from b towards a and c. At this moment conduction from phase B is terminating and commencing from phase A. Thus current flows from phase A through 1 and 1′ for a period corresponding to $(2\pi/3 - \pi/3)$ radians, i.e. 60°. At $2\pi/3$ the voltage of phase C commences to exceed that of phase A, and thus phase C delivers current through 1 and 2′. After conduction for a period corresponding to 60°, phase B takes over and delivers current through 2 and 2′. Continuing in this manner, it will be found that the following summarizes the conditions over a cycle

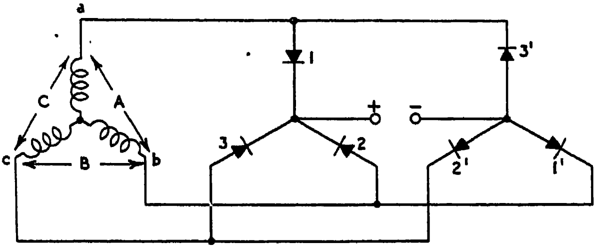


FIG. 14–10

A over from B through 1 and 1′
C ,, ,, A ,, 1 ,, 2′
B ,, ,, C ,, 2 ,, 2′
A ,, ,, B ,, 2 ,, 3′
C ,, ,, A ,, 3 ,, 3′
B ,, ,, C ,, 3 ,, 1′

As with previous circuits $m = 6$ and thus $V_s = 0{\cdot}74E_0$, V_s in this case referring to the voltage between lines. Current commences to flow in phases a and b for $\theta = \pi/3$, continuing until $\theta = 2\pi/3$. This is immediately followed by a similar pulse through a and c. Thus, each anode (or rectifying element) carries two pulses of current per cycle and the r.m.s. value of the anode current is $\sqrt{2}$ times that of the six-phase circuits previously discussed. Hence

$$I_a = \sqrt{2} \times 0{\cdot}41I = 0{\cdot}58I$$

Now each secondary phase carries four pulses of current per cycle, each of which is equal to that of the anode currents.

Thus $$I_s = \sqrt{2} \times 0{\cdot}58I = 0{\cdot}82I$$

Due to the three-phase character of the secondary winding, the primary current is given by

$$I_p = (N_s/N_p)I_s = 0{\cdot}82(N_s/N_p)I$$

Transformer Secondary Rating

For the star-connected secondary shown, the rating is

$$3 \cdot \frac{V_s}{\sqrt{3}} I_s = 3 \times \frac{0{\cdot}74E_0}{\sqrt{3}} \times 0{\cdot}82I = 1{\cdot}05E_0I$$

the transformer primary also having the same rating. As two phases are always simultaneously carrying current, it follows that a star-connected primary winding may be employed.

The circuit of Fig. 14–10 may also be employed with a delta-connected secondary winding. In this case the vector sum of the e.m.f.s of any two phases is equal to that of the remaining phase. Also if the impedances of all phases are equal, the impedance of a pair of phases must be twice that of one. Thus, with a delta-connected winding, if a load is connected to one phase, two-thirds of the load current is supplied by this phase and the remaining third by the other pair. It follows that while the voltage of phase A, Fig. 14–10, is higher than either of the voltages of B or C, A is contributing two-thirds of the current to the line and B and C one-third. Thus, if I_{Lm} is the maximum value of the secondary line current, the r.m.s. value of the phase current is

$$\sqrt{\frac{1}{\pi}\left[2\int_{\pi/3}^{2\pi/3} \frac{I_{Lm}{}^2}{3^2} \sin^2 \theta d\theta + (\tfrac{2}{3} I_{Lm})^2 \int_{\pi/3}^{2\pi/3} \sin^2 \theta d\theta\right]}$$

$$= \sqrt{\frac{6I_{Lm}{}^2}{9\pi} \int_{\pi/3}^{2\pi/3} \sin^2 \theta d\theta}$$

$$= I_{Lm}\sqrt{\frac{1}{9} + \frac{\sqrt{3}}{6\pi}} = 0{\cdot}45I_{Lm}$$

For six-phase rectification $I_{Lm} = 1{\cdot}047I$ and, therefore, we have

$$I_s = 0{\cdot}472I$$

The secondary rating with a delta connexion is

$$3 \times 0{\cdot}472I \times 0{\cdot}74E_0 = 1{\cdot}05E_0I$$

which is the same as when a star-connexion is employed. As with a star-connected secondary, the secondary line and rectifier currents are respectively equal to $0{\cdot}82I$ and $0{\cdot}58I$.

Either a star- or delta-connected primary winding may be employed in connexion with a delta-connexion on the secondary side.

A.C. Supply Loading

With a star-connected primary and secondary the primary line current is the same as the phase current, i.e. $0{\cdot}82(N_s/N_p)I$. Hence the supply loading is

$$\begin{aligned}\sqrt{3}V_LI_L &= \sqrt{3}\,.\,\sqrt{3}V_p \times 0{\cdot}82(N_s/N_p)I\\ &= 3(N_p/N_s)\frac{V_s}{\sqrt{3}} \times 0{\cdot}82(N_s/N_p)I\\ &= \sqrt{3}V_s \times 0{\cdot}82I\\ &= \sqrt{3} \times 0{\cdot}74E_0 \times 0{\cdot}82I\\ &= 1{\cdot}05E_0I\end{aligned}$$

If the secondary is connected in delta and the primary in star, then the rating is

$$\begin{aligned}\sqrt{3}V_LI_L &= \sqrt{3}\,.\,\sqrt{3}V_p \times 0{\cdot}472(N_s/N_p)I\\ &= 3(N_p/N_s)V_s \times 0{\cdot}472(N_s/N_p)I\\ &= 3 \times 0{\cdot}74E_0 \times 0{\cdot}472I\\ &= 1{\cdot}05E_0I\end{aligned}$$

Where a delta-connected primary is used with a star-connected secondary the primary line current is $\sqrt{3} \times 0{\cdot}82(N_s/N_p)I = 1{\cdot}42(N_s/N_p)I$. The supply loading is, then,

$$\begin{aligned}\sqrt{3} \times (N_p/N_s)\frac{0{\cdot}74E_0}{\sqrt{3}} &\times 1{\cdot}42(N_s/N_p)I\\ &= 1{\cdot}05E_0I\end{aligned}$$

The same result is obtained if a delta-delta connexion is employed.

The power factor is, of course, $1/1{\cdot}05 = 0{\cdot}955$ and the ripple is the same as with the six-phase circuits previously discussed.

Comparing the six-phase bridge circuit with the two foregoing circuits, we observe that in the former case the secondary and primary ratings and supply loadings are the same; i.e. $1{\cdot}05E_0I$. This is considerably lower than the primary ratings of the latter circuits and very considerably lower than their secondary ratings. On the other hand, the specific anode loading of the bridge circuit is higher than that of the other two in the ratio of 0·58 to 0·41, i.e. it is 41 per cent greater. In this respect the bridge circuit is at something of a disadvantage, for it may necessitate a larger rectifier or, in the case of metal rectifiers, more discs in parallel for a given output than would be the case with the previously described circuits. However, the simplicity and low rating of the transformer of the bridge circuit tend to offset this disadvantage.

Considering the value of the anode current obtained in practice, due to the secondary windings carrying the current of two consecutive anodes, m in (13–25) must be taken as 3. Hence

$$I_a = 0{\cdot}58I\sqrt{1-3f(u)}$$

from which it will be observed that the reduction in anode current is not as low as for previous circuits for the same value of u.

A point of particular interest of the bridge circuit is the value of the peak inverse voltage. Compared with that of the circuits of Figs. 14–8 and 14–9, the value is only one-half in this case. Where metal rectifiers are employed this is a matter of great importance, for the necessary number of discs in series per phase is equal to the peak inverse voltage divided by the maximum permissible volts per disc. Hence on this score alone, for a given output voltage, the bridge circuit requires only one-half the number of discs needed by either of the other six-phase circuits.

Taking into consideration the effect of transformer reactance and assuming a 1 : 1 transformation ratio, the secondary current is

$$I_s = 0{\cdot}82I\sqrt{1-3f(u)} \quad \text{(star-connected)}$$

$$I_s = 0{\cdot}472I\sqrt{1-3f(u)} \quad \text{(delta-connected)}$$

similar values obtaining for the primary winding. From (13–23), and assuming a delta-star connexion,

$$E = E_0 - \left[\frac{m}{\pi}\cdot\frac{IX}{2}\right]$$

$$= E_0 - \left[\frac{3}{\pi}\cdot\frac{IX}{2}\right]$$

As with the triple-star connexion we must take for the reactance $X/3$. Thus

$$E = E_0 - \left[\frac{IX}{2\pi}\right]$$

Expressing I in terms of I_p

$$I = \frac{1{\cdot}22 I_p}{(N_s/N_p)}\left[\frac{1}{\sqrt{1-3f(u)}}\right]$$

and, as $N_s = N_p/\sqrt{3}$,

$$I = 2{\cdot}1 I_p\left[\frac{1}{\sqrt{1-3f(u)}}\right]$$

Thus
$$E = \left(E_0 - \left[\frac{0{\cdot}335 I_p X_p}{\sqrt{1-3f(u)}}\right]\right)$$
$$= 1{\cdot}35\left(V_p - \left[\frac{I_p X_p}{4\sqrt{1-3f(u)}}\right]\right)$$

which indicates that the bridge circuit possesses particularly low regulation, actually 50 per cent of that of the triple-star circuit.

The primary and secondary ratings and supply loading are given by

$$1{\cdot}05EI\left[\frac{\sqrt{1-3f(u)}}{\cos^2\dfrac{u}{2}}\right]$$

Phase Equalizing

In the rectifier circuits so far dealt with, it is seen that, provided the interval of overlap be ignored, each rectifier anode or element carries the d.c. load current for $2\pi/m$ of a cycle only. This leads to poor utilization of the rectifier phases and results in transformers which are large and uneconomical. In view of this, circuits have been devised which enable two, three, or four anodes to carry current simultaneously and which, moreover, allow each anode to carry current for a longer period than the multi-phase circuits previously discussed.

Now, normally, an anode only carries current while the voltage of the phase to which it is connected is higher than the rest. We have seen, however, that during overlap two anodes may simultaneously carry current and that while they do so their phase voltages are equal. Hence, in order that two or more anodes may simultaneously carry current, it is essential that the phase voltages shall be of identical value during the common period of current conduction. This is effected in practice by means of what is termed phase equalizing, i.e. by electromagnetically coupling phases or phase groups in such a manner that two or more phase voltages are simultaneously equal. Reference to Fig. 14–11 will indicate the manner in which this is carried out for most large six-phase installations. It will be noted that the phases are grouped into two groups of three, star-connected, with a reactor connecting the star points of each group. The reactor is known as an interphase reactor or transformer, and it will be noted that its electrical centre forms the d.c. negative terminal. As (with the circuit shown) two anodes simultaneously conduct, two phases are always connected via the arcs and cathode pool. Thus, if the arc drop is ignored, the e.m.f. impressed on the interphase reactor is the difference between those of the two-phase groups, and hence half this difference will either be added to or subtracted from each group, thus equalizing the phase voltages and permitting two phases, one of each group, simultaneously to carry current. It will be noted that each phase group contributes one-half of the d.c. load current, each part of this passing through the reactor, the windings of this being so arranged that the m.m.f.s due to the d.c. current oppose each other, resulting in an absence of core magnetization.

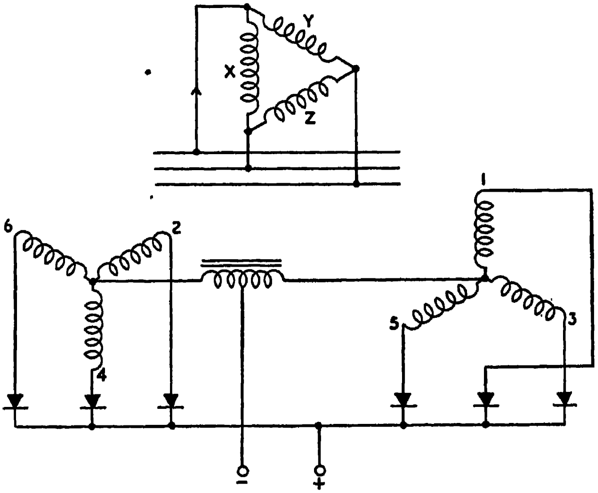


FIG. 14–11

Fig. 14–12 shows the instantaneous voltages and their resultants acting in the circuit of Fig. 14–11. It will be seen that the instantaneous d.c. voltage is given by

$$\frac{\sqrt{2}V_s \sin \theta + \sqrt{2}V_s \sin (\theta - \pi/3)}{2} = \sqrt{\frac{3}{2}}\, V_s \sin (\theta - \pi/6)$$

the mean value being

$$\frac{3}{\pi}\sqrt{\frac{3}{2}}\, V_s \int_{\pi/2}^{5\pi/6} \sin (\theta - \pi/6) d\theta \qquad . \qquad . \quad (14\text{–}10)$$

where the limits refer to phase 1.

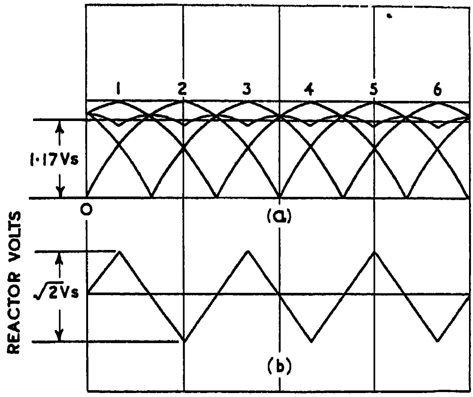


FIG. 14–12

Integrating (14–10)

$$E_0 = \frac{3\sqrt{3}}{\sqrt{2}\pi} V_s = 1{\cdot}17 V_s$$

and $$V_s = 0{\cdot}86 E_0$$

which is the same as for a simple three-phase system. It will be noted that the triple-frequency fluctuations of each phase group combine to produce a smaller fluctuation of sextuple frequency, as in the six-phase circuit of Fig. 14–8. The amplitude of the fluctuation is

$$\sqrt{2}V_s \sin \pi/3 - (\sqrt{2}V_s \sin \pi/2 + \sqrt{2}V_s \sin \pi/6)/2$$
$$= \sqrt{\frac{3}{2}}\, V_s \left(1 - \frac{\sqrt{3}}{2}\right)$$
$$= 0{\cdot}162 V_s \text{ or } 0{\cdot}138 E_0$$

which is the same as for a simple six-phase circuit. From the foregoing it will be appreciated that the circuit of Fig. 14–11 behaves in one respect as a simple three-phase rectifier with a relatively high transformer utility factor, and in another as a six-phase rectifier with its relatively low d.c. voltage ripple.

The instantaneous voltage across each *half* of the interphase reactor is

$$\frac{\sqrt{2}V_s \sin \theta - \sqrt{2}V_s \sin (\theta - \pi/3)}{2} = \frac{V_s}{\sqrt{2}} \cos (\theta - \pi/6)$$

one maximum of this occurring for $\theta = \pi/2$, referred to phase 1, Fig. 14–12.

The r.m.s. of this voltage is

$$\sqrt{\frac{V_s^2}{2} \cdot \frac{6}{\pi} \int_{\pi/2}^{2\pi/3} \cos^2 (\theta - \pi/6) d\theta} \quad . \quad . \quad (14\text{–}11)$$

the limits being referred to phase 1.

Integrating, (14–11) becomes

$$\sqrt{\frac{2\pi - 3\sqrt{3}}{8\pi}}\, V_s$$

From (13–7), $V_s = \sqrt{2}\pi E_0/3\sqrt{3}$ and thus the reactor voltage may be written

$$\frac{\sqrt{2}\pi}{3\sqrt{3}} \sqrt{\frac{2\pi - 3\sqrt{3}}{8\pi}}\, E_0 = 0{\cdot}177E_0 \quad . \quad . \quad (14\text{–}12)$$

Reference to Fig. 14–12 (*b*) shows that the waveform of the reactor voltage is approximately triangular, and that the maximum voltage across the *entire* reactor winding is $V_s/\sqrt{2}$ volts.

Currents and Ratings

As each three-phase group contributes one-half of the d.c. output current, the phase current is one-half of that occurring with simple three-phase working, i.e. $0{\cdot}29I$. Each primary phase must supply two secondary phases in turn during a cycle, and, therefore, the primary phase current is

$$\sqrt{2}(N_s/N_p)I_s = 0{\cdot}41(N_s/N_p)I$$

We may note that as two secondary phases are simultaneously in operation, so also must be two primary phases. Thus, in this case, a star-connected primary winding may be employed.

Transformer Secondary Rating

The secondary rating is

$$6V_sI_s = 6 \times 0{\cdot}86E_0 \times 0{\cdot}29I = 1{\cdot}5E_0I$$

Transformer Primary Rating

The primary rating is

$$\begin{aligned} 3V_pI_p &= 3 \times 0{\cdot}86(N_p/N_s)E_0 \times 0{\cdot}41(N_s/N_p)I \\ &= 3 \times 0{\cdot}86E_0 \times 0{\cdot}41I = 1{\cdot}05E_0I \end{aligned}$$

The mean rating is thus $1{\cdot}27E_0I$.

INTERPHASE REACTOR RATING

For the purpose of determining the reactor rating it may be regarded as a 1 : 1 transformer, each winding carrying a d.c. current $I/2$, plus a negligible magnetizing current. The voltage across each winding is $0{\cdot}177E_0$, and thus the rating is

$$\frac{0{\cdot}177E_0I}{2} = 0{\cdot}088E_0I$$

Hence the mean rating of the transformer unit as a whole is $1{\cdot}36E_0I$. It may be remembered, however, that the frequency of the reactor impressed e.m.f. is three times that of the supply frequency, and this generally permits a reduction in the size of the reactor compared with mains frequency working.

A.C. Supply Loading

On the assumption that a star-connected primary is employed, the line current is equal to the phase current, i.e.

$$I_L = I_p = 0{\cdot}41(N_s/N_p)I$$

The supply loading is then

$$\begin{aligned} \sqrt{3}V_LI_L &= \sqrt{3}\,.\,\sqrt{3}V_p \times 0{\cdot}41(N_s/N_p)I \\ &= 3(N_p/N_s)V_s \times 0{\cdot}41(N_s/N_p)I \\ &= 3 \times 0{\cdot}86E_0 \times 0{\cdot}41I \\ &= 1{\cdot}05E_0I \end{aligned}$$

Assuming a delta-connected primary is employed, the line current may be determined in the following manner. Referring to Fig. 14–11, when anodes 1 and 6 are conducting, the line connected to the top of phase X carries the primary currents for phases X and Y for one-sixth of a cycle, these two currents, of course, being 60°

out of phase with each other. Next, anodes 1 and 2 conduct, during which time the same line carries the current of the X phase only for one-sixth of a cycle. Following this, anodes 2 and 3 conduct, during which time the line carries the current of the Y phase for a further one-sixth of a cycle, but in the reverse direction to formerly. Hence the r.m.s. value of the line current is given by

$$\sqrt{\frac{1}{\pi}\left(\frac{N_s}{N_p}I_m\right)^2\left\{\int_{\pi/2}^{5\pi/6}[\sin\theta+\sin(\theta-\pi/3)]^2d\theta+\int_{\pi/2}^{5\pi/6}\sin^2\theta d\theta+\int_{\pi/6}^{\pi/2}\sin^2\theta d\theta\right\}}$$

where the limits of the first integral refer to anode 6, integration being taken from the moment anode 1 starts to conduct. Thus the first integral refers to the one-sixth of a cycle, while anodes 1 and 6 are conducting, the second to when anode 1 is conducting for a further one-sixth of a cycle, anode 6 having terminated, and the third integral to when anode 3 is conducting. Integrating, we have

$$\begin{aligned}I_L&=\frac{N_s}{N_p}I_m\sqrt{\frac{1}{\pi}\left[\left(\frac{\pi}{2}+\frac{3\sqrt{3}}{4}\right)+\left(\frac{\pi}{6}+\frac{\sqrt{3}}{8}\right)+\left(\frac{\pi}{6}+\frac{\sqrt{3}}{8}\right)\right]}\\&=\frac{N_s}{N_p}I_m\sqrt{\frac{5}{6}+\frac{\sqrt{3}}{\pi}}\\&=1{\cdot}18(N_s/N_p)I_m\end{aligned}$$

We have already seen that for a simple three-phase circuit $I_m = 1{\cdot}21I$. Hence for the circuit of Fig. 14–11, $I_m = 0{\cdot}6I$ as each three-phase group provides one-half of the d.c. output current. The line current is then given by

$$I_L = 1{\cdot}18(N_s/N_p)0{\cdot}6I = 0{\cdot}708(N_s/N_p)I$$

In this case the supply loading is

$$\begin{aligned}\sqrt{3}V_LI_L&=\sqrt{3}V_pI_L=\sqrt{3}(N_p/N_s)V_sI_L\\&=\sqrt{3}(N_p/N_s)\times 0{\cdot}86E_0\times 0{\cdot}708(N_s/N_p)I\\&=\sqrt{3}\times 0{\cdot}86E_0\times 0{\cdot}708I\\&=1{\cdot}05E_0I\end{aligned}$$

as before.

Contrasting this circuit with those of Figs. 14–8 and 14–9, we note that the rating of the transformer shows a reduction of 19 per cent on that of Fig. 14–8 and 11·5 per cent on that of Fig. 14–9. If the interphase reactor rating is included, then the respective percentage reductions are 13·3 and 5·2. It may also be noted that compared with the circuits of Figs. 14–8 and 14–9, the phase-equalizing circuit reduces the anode loading in the ratio of 0·29 to 0·41. This, of course, permits an increased output from a given rectifier or a reduction in the rectifier dimensions for a given output. As with simple six-phase working the power factor is 0·95.

An effect which may be experienced with the last-described circuit is a reversion to simple six-phase working on very light loads, with a consequent increase in d.c. voltage from $1{\cdot}17V_s$ to $1{\cdot}33V_s$. This is caused by the load current becoming less than that needed to magnetize the reactor. If i is the r.m.s. value of the magnetizing current, then the d.c. load current at which the foregoing effect occurs is $2\sqrt{2}i$ amp., this being known as the transition current. Where this phenomenon is objected to, it may be obviated by the connexion of a small permanent load to the output terminals.

Considering the effects of transformer reactance

$$I_a = 0{\cdot}29I[\sqrt{1 - 3f(u)}]$$

$$I_p = 0{\cdot}41I[\sqrt{1 - 3f(u)}]$$

The angle of overlap in the above equations is determined from (13–22), but I in that equation must be replaced by $I/2$. Similarly in (13–23), I must be replaced by $I/2$. Hence

$$E = E_0 - \left[\frac{3}{4\pi} IX\right]$$

Expressing I in terms of I_p and assuming a 1 : 1 transformation ratio

$$I = 2{\cdot}44I_p\left[\frac{1}{\sqrt{1 - 3f(u)}}\right]$$

and

$$E = E_0 - \left[\frac{2{\cdot}44I_pX_p}{4\pi\sqrt{1 - 3f(u)}}\right]$$

$$= 1{\cdot}17\left(V_p - \left[\frac{I_pX_p}{2\sqrt{1 - 3f(u)}}\right]\right)$$

The secondary rating is

$$1{\cdot}5EI\left[\frac{\sqrt{1 - 3f(u)}}{\cos^2\frac{u}{2}}\right]$$

and the primary rating and supply loading

$$1{\cdot}05EI\left[\frac{\sqrt{1 - 3f(u)}}{\cos^2\frac{u}{2}}\right]$$

Twelve-phase Rectification

It is evident that the greater the number of phases employed, the smaller the d.c. voltage fluctuation, or ripple, becomes. It will also have been noted that as the phase number is increased so also is the rectifier power factor. Hence it follows that by employing twelve-phase rectification a further reduction in ripple and increase in power factor may be obtained, compared with those obtainable with the circuits previously described. Disadvantages associated with twelve-phase rectification are poor utility factors and complicated and costly transformer windings. In fact, it may be held that, where a particularly smooth output is required, it is preferable to employ six-phase rectification with smoothing equipment rather than the somewhat complicated twelve-phase rectifier.

The Twelve-phase Fork Circuit

This circuit is a development of the six-phase fork circuit of Fig. 14–9 and is shown by Fig. 14–13. The complicated character of the secondary windings is at once apparent, there being no fewer than eighteen secondary windings accommodated on the three transformer limbs. It will be noted that the phase voltage V_s is produced by the combination of voltages V_2 and V_3. From the sine rule

$$\frac{V_s}{\sin 120^\circ} = \frac{V_3}{\sin 15^\circ} = \frac{V_2}{\sin 45^\circ}$$

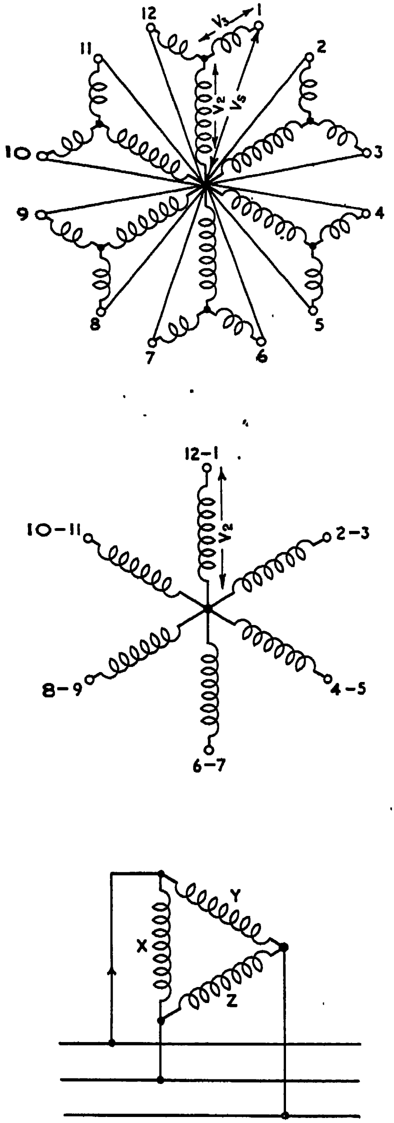


FIG. 14–13

and thus we have

$$V_3 = 0{\cdot}3V_s$$
$$V_2 = 0{\cdot}816V_s$$

As $m = 12$, we have

$$V_s = \frac{E_0}{12\sqrt{2}\sin \pi/12} = 0{\cdot}715E_0$$

$$I_a = \frac{\pi\sqrt{\frac{1}{2\pi}\left(\frac{\pi}{12} + \frac{1}{2}\sin\frac{2\pi}{12}\right)}}{12\sin \pi/12} = 0{\cdot}29I$$

$$I_m = \frac{\pi}{12\sin \pi/12}I = 1{\cdot}015I \simeq I \text{ approx.}$$

Now the windings of the centre star each supply two anodes in succession per cycle and hence

$$I_s = \sqrt{2}I_a = 0{\cdot}41I$$

In order to determine the primary-phase current we may note that phase X, for example, supplies, in succession, per cycle, the windings 12–1, 2, 5, 6–7, 8, and 11. Thus, remembering that 12–1 and 6–7 each conduct twice per cycle, the primary-phase current is given by

$$\sqrt{\frac{1}{2\pi}\cdot 4\left(\frac{N_s}{N_p}\right)^2 I_m{}^2\int_{5\pi/12}^{7\pi/12}\sin^2\theta d\theta + 4\left(\frac{N_{s1}}{N_p}\right)^2 I_m{}^2\int_{5\pi/12}^{7\pi/12}\sin^2\theta d\theta}$$

$$= \sqrt{\frac{1}{2\pi}\left(\frac{N_s}{N_p}\right)^2 I_m{}^2\left(\frac{\pi+3}{3}\right) + \left(\frac{N_{s1}}{N_p}\right)^2 I_m{}^2\left(\frac{\pi+3}{3}\right)}$$

$$= \left(\frac{N_s}{N_p}\right)I\sqrt{\frac{1}{2\pi}\cdot 1{\cdot}015^2\left(\frac{\pi+3}{3}\right)\left[1+\left(\frac{0{\cdot}3}{0{\cdot}816}\right)^2\right]}$$

$$= 0{\cdot}61(N_s/N_p)I, \text{ or } 0{\cdot}5(V_s/V_p)I$$

Transformer Secondary Rating

The transformer secondary rating is

$$\begin{aligned} 12V_3I_a + 6V_2I_s &= 12 \times 0{\cdot}3 \times 0{\cdot}72E_0 \times 0{\cdot}29I \\ &\quad + 6 \times 0{\cdot}816 \times 0{\cdot}72E_0 \times 0{\cdot}41I \\ &= 0{\cdot}75E_0I + 1{\cdot}45E_0I \\ &= 2{\cdot}2E_0I \end{aligned}$$

Transformer Primary Rating

The primary rating is

$$\begin{aligned} 3V_pI_p &= 3(N_p/N_s)V_2 \times 0{\cdot}61(N_s/N_p)I \\ &= 3 \times 0{\cdot}816V_s \times 0{\cdot}61I \\ &= 3 \times 0{\cdot}816 \times 0{\cdot}72E_0 \times 0{\cdot}61I \\ &= 1{\cdot}08E_0I \end{aligned}$$

The mean rating is thus $1{\cdot}64E_0I$.

A.C. Supply Loading

In order to determine the line current, we may note that, when anode 1 is conducting, the line connected to the top of phase X carries the current for this phase for one-twelfth of a cycle, the instantaneous value of the current being $(N_s/N_p)I_m \sin\theta$. When the arc shifts to anode 2 the line current becomes $(N_{s1}/N_p)I_m \sin\theta$ and has the same value, but reversed in direction, when anode 3 conducts. For anode 4 the line current is $(N_s/N_p)I_m \sin\theta$, while for anode 5 the value is $[(N_s/N_p)I_m + (N_{s1}/N_p)I_m] \sin\theta$. Continuing in this fashion round the secondary circuit, the r.m.s. value of the line current will be seen to be,

$$\sqrt{\frac{1}{2\pi}\left\{4(N_s/N_p)^2I_m{}^2\int_{5\pi/12}^{7\pi/12}\sin^2\theta d\theta + 4(N_{s1}/N_p)^2I_m{}^2\int_{5\pi/12}^{7\pi/12}\sin^2\theta d\theta \right.}$$

$$\left. + 4\int_{5\pi/12}^{7\pi/12}[(N_s/N_p)I_m\sin\theta + (N_{s1}/N_p)I_m\sin\theta]^2d\theta\right\}$$

$$= \sqrt{\frac{1}{2\pi}\left\{4{\cdot}54(N_s/N_p)^2I_m{}^2\int_{5\pi/12}^{7\pi/12}\sin^2\theta d\theta + 7{\cdot}53(N_s/N_p)^2I_m{}^2\int_{5\pi/12}^{7\pi/12}\sin^2\theta d\theta\right\}}$$

$$= \sqrt{\frac{12{\cdot}07}{2\pi}(N_s/N_p)^2I_m{}^2\int_{5\pi/12}^{7\pi/12}\sin^2\theta d\theta}$$

$$= \sqrt{\frac{12{\cdot}07}{2\pi}(N_s/N_p)^2I_m{}^2\left(\frac{\pi+3}{12}\right)}$$

$$= (N_s/N_p)I, \text{ or } 0{\cdot}816(V_s/V_p)I$$

The supply loading is

$$\begin{aligned}\sqrt{3}V_L I_L &= \sqrt{3}V_p \times (N_s/N_p)I \\ &= \sqrt{3}(N_p/N_s)V_2 \times (N_s/N_p)I \\ &= \sqrt{3} \times 0{\cdot}816 \times 0{\cdot}715E_0 \times I \\ &= 1{\cdot}01E_0 I\end{aligned}$$

It will be noted that the algebraical sum of the currents in the primary is not equal to zero and, therefore, only a delta-connected winding may be employed.

Power Factor

The power factor is given by

$$\frac{E_0 I}{1{\cdot}01E_0 I} = 0{\cdot}99$$

Voltage Fluctuation

In this case the output voltage fluctuates between $\sqrt{2}V_s$ and $\sqrt{2}V_s \sin 7\pi/12$. The amplitude of the fluctuation is

$$\begin{aligned}&\sqrt{2}V_s - \sqrt{2}V_s \sin 7\pi/12 \\ &\sqrt{2}(1 - 0{\cdot}9659)V_s = 0{\cdot}048V_s \\ &\qquad\qquad\qquad\qquad\; = 0{\cdot}034E_0\end{aligned}$$

Maximum Inverse Voltage

Ignoring the drop in the conducting anode, the non-conducting anodes must each withstand $2\sqrt{2}V_s$.

THE QUADRUPLE THREE-PHASE CIRCUIT

The poor utility factor and high transformer rating of the foregoing circuit has, as with six-phase rectification, led to the introduction of circuits employing the principle of phase-equalizing. Several such circuits exist but, here, only what is probably the best and most frequently employed circuit will be discussed. This circuit is shown by Fig. 14–14, and, like that of Fig. 14–13, possesses a complicated secondary winding which, virtually, consists of four three-phase groups coupled by three interphase reactors. As with the circuit of Fig. 14–13, the phase voltage V_s is produced by the combination of the voltages V_2 and V_3. Hence, as before,

$$\begin{aligned}V_3 &= 0{\cdot}3V_s \\ V_2 &= 0{\cdot}816V_s\end{aligned}$$

Each of the three-phase groups of Fig. 14–14 supplies one-quarter of the d.c. current (four anodes, one from each group, being simultaneously in operation), each anode carrying current for one-third of a cycle. A six-phase group, such as 1, 2, functions in an exactly similar manner to that of Fig. 14–11 and hence $V_s = 0{\cdot}86E_0$. It will be noted also that as each three-phase group contributes one-quarter of the d.c. output current, the secondary phase and anode currents are one-half of those of Fig. 14–11, i.e. $0{\cdot}145I$.

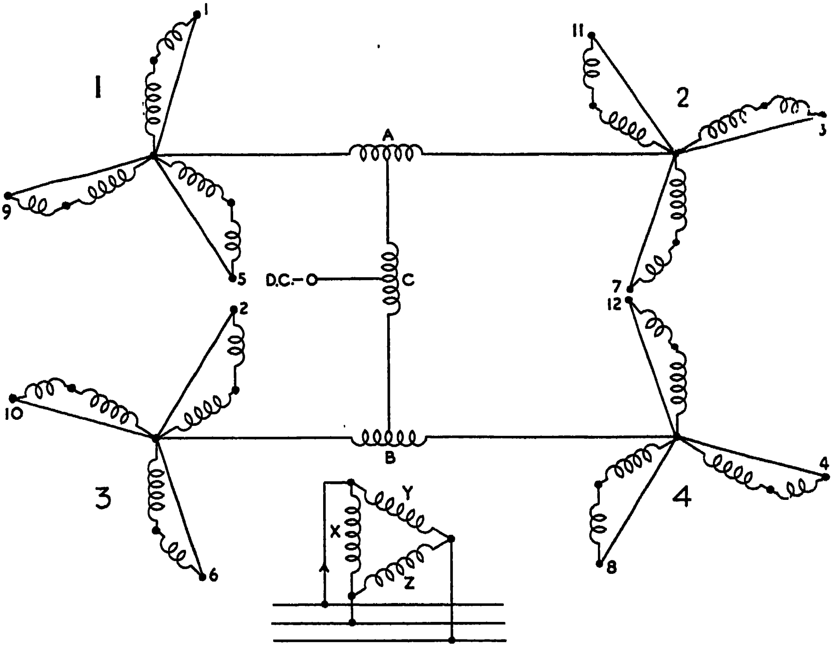


Fig. 14–14

In order to determine the primary current we note that as anode 12 finishes conduction (Fig. 14–15), anode 4 is just commencing. The four anodes which are at any instant conducting are, of course, those the potentials of which occur in sequence, such as, for example, 10, 11, 12, 1, — 11, 12, 1, 2, — 12, 1, 2, 3, — . The value of the primary current may, of course, be determined in a similar manner to that employed for finding the line current of Fig. 14–13. However, due to the complicated nature of the transformer windings in the present case, this method leads to lengthy and tedious integration of an excessive number of terms. To avoid this difficulty, the simplifying assumption will be made that the anode currents are of rectangular waveform of amplitude $0{\cdot}25I$. This corresponds to when a cathode choke is employed or an inductive load is supplied. Under these circumstances the r.m.s. value of the anode current is

$$I_a = \sqrt{(0{\cdot}25I)^2 \times \frac{1}{2\pi}\int_{\frac{m-2}{2m}\pi}^{\frac{m+2}{2m}\pi} d\theta}$$

$$= \frac{0{\cdot}25I}{\sqrt{m}} = \frac{0{\cdot}25I}{\sqrt{3}} = 0{\cdot}144I$$

which differs by a negligible amount from the figure already derived by a more rigorous method. It may be here recalled that it has already been shown (page 443) that for $m \geq 3$, $I_a = I/\sqrt{m}$, approximately.

Considering the moment when anodes 1, 2, 3, and 4 are conducting, the current in the primary phase X is

$$0{\cdot}25(N_s/N_p)I + 0{\cdot}25(N_{s1}/N_p)I = 0{\cdot}34(N_s/N_p)I$$

As conduction through anode 1 terminates, it commences through anode 5. The X-phase current is now that necessary to supply the outer windings of anodes 2 and 5. This current is, therefore,

$$0{\cdot}25(N_{s1}/N_p)I - 0{\cdot}25(N_{s1}/N_p)I = 0$$

$\pi/6$ radians later anodes 3, 4, 5, and 6 are conducting and the X-phase is supplying current for the outer winding of anode 5 and the inner winding of anode 6. The current now is

$$-\,0{\cdot}25(N_{s1}/N_p)I - 0{\cdot}25(N_s/N_p)I = -\,0{\cdot}34(N_s/N_p)I$$

Continuing in this manner for a complete cycle, the r.m.s. value of the primary phase current is given by

$$\begin{aligned}I_p = (N_s/N_p)I\sqrt{\frac{1}{12}}\,[&(0{\cdot}34)^2 + (0)^2 + (-\,0{\cdot}34)^2 + (-\,0{\cdot}59)^2 \\ &+ (-\,0{\cdot}683)^2 + (-\,0{\cdot}59)^2 + (-\,0{\cdot}34)^2 + (0)^2 \\ &+ (-\,0{\cdot}34)^2 + (0{\cdot}59)^2 + (0{\cdot}683)^2 + (0{\cdot}59)^2] \\ = 0{\cdot}484&(N_s/N_p)I \text{ or } 0{\cdot}395(V_s/V_p)I\end{aligned}$$

Transformer Secondary Rating

The transformer secondary rating is

$$\begin{aligned}12V_3I_a + 12V_2I_s &= 12V_sI_s(0{\cdot}816 + 0{\cdot}3) \\ &= 12 \times 0{\cdot}86E_0 \times 0{\cdot}145I \times 1{\cdot}116 = 1{\cdot}67E_0I\end{aligned}$$

The primary rating is

$$\begin{aligned} 3V_pI_p &= 3(N_p/N_s)V_2 \times 0{\cdot}484(N_s/N_p)I \\ &= 3 \times 0{\cdot}816 \times 0{\cdot}86E_0 \times 0{\cdot}484I \\ &= 1{\cdot}01E_0I \end{aligned}$$

The mean rating is thus $1{\cdot}34E_0I$.

Interphase Reactor Ratings

The reactors A and B connecting groups 1, 2, and 3, 4, function in an exactly similar manner to that of Fig. 14–11. Their windings, however, carry only one-quarter of the d.c. output current and hence their rating is $0{\cdot}044E_0I$. The function of the reactor C is to couple the two six-phase groups. In consequence, each of its windings will carry one-half of the d.c. load current, while the voltage across this reactor will be the difference of the instantaneous e.m.f.s across each six-phase group. Reference to Fig. 14–15 will show the twelve phase voltages and combining 1 and 3, 3 and 5, etc., and also 2 and 4, 4 and 6, etc., we obtain the instantaneous voltages across each six-phase group. The difference between these voltages is shown by the smaller triangular wave which, it is to be noted, is of sextuple frequency. Its amplitude is equal to the difference between the maximum and minimum values of the voltage of either six-phase group, i.e. $0{\cdot}162V_s$ or $0{\cdot}138E_0$. As the r.m.s. value of a triangular wave of the form shown is $1/\sqrt{3}$ times the peak value, the r.m.s. voltage across C is $0{\cdot}138E_0/\sqrt{3}$. One-half of this voltage is, of course, across each half of the reactor and thus its rating is

$$\frac{0{\cdot}138E_0}{2\sqrt{3}} \times \frac{I}{2} = 0{\cdot}02E_0I$$

The total rating of the three reactors is

$$2 \times 0{\cdot}044E_0I + 0{\cdot}02E_0I = 0{\cdot}108E_0I$$

and thus the total transformer rating amounts to

$$1{\cdot}34E_0I + 0{\cdot}108E_0I = 1{\cdot}448E_0I$$

A.C. Supply Loading

A consideration of the circuit of Fig. 14–14 will show that two primary phases are always simultaneously in operation, carrying identical currents. For example, when anodes 1, 2, 3, and 4 are conducting, the current loadings in phases X and Y are identical. It therefore follows that a star-connected primary may be employed.

In this case the line current is equal to the phase current, i.e. $0{\cdot}484(N_s/N_p)I$. The supply loading is then

$$\begin{aligned}\sqrt{3}V_L I_L &= 3(N_p/N_s)V_2 \times 0{\cdot}484(N_s/N_p)I\\ &= 3 \times 0{\cdot}816 \times 0{\cdot}86E_0 \times 0{\cdot}484I\\ &= 1{\cdot}01E_0 I\end{aligned}$$

In the event of a delta-connected primary being employed, the line current may be derived in a similar manner to that used for determining the primary-phase currents. Thus, when anodes 1, 2, 3, and 4 are in operation, the line connected to the top of phase X must carry—

$$\begin{array}{cccc} \textit{Anode } 1 & \textit{Anode } 2 & \textit{Anode } 3 & \textit{Anode } 4\\ 0{\cdot}25(N_s/N_p)I & +\ 0{\cdot}25(N_{s1}/N_p)I & -\ 0{\cdot}25(N_{s1}/N_p)I & -\ 0{\cdot}25(N_s/N_p)I \end{array}$$
$$= 0$$

$\pi/6$ radians later the line current is

$$\begin{array}{cccc} \textit{Anode } 2 & \textit{Anode } 3 & \textit{Anode } 4 & \textit{Anode } 5\\ 0{\cdot}25(N_{s1}/N_p)I & -\ 0{\cdot}25(N_{s1}/N_p)I & -\ 0{\cdot}25(N_s/N_p)I & -\ 0{\cdot}25(N_s/N_p)I\\ & & & -\ 0{\cdot}25(N_{s1}/N_p)I \end{array}$$
$$= -\,0{\cdot}592(N_s/N_p)I$$

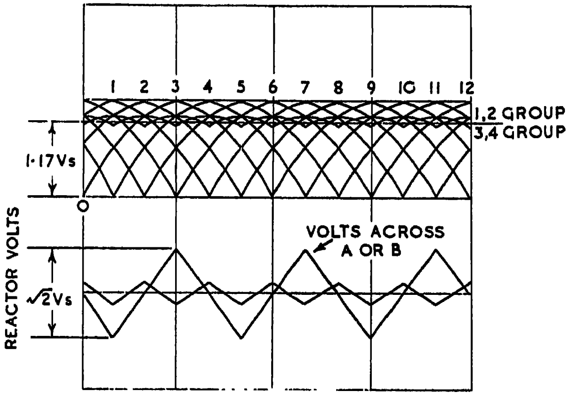


FIG. 14–15

Continuing in this manner for a complete cycle, the r.m.s. value of the line current is given by

$$\begin{aligned}I_L = (N_s/N_p)I\sqrt{\tfrac{1}{12}}\,[&(0)^2 + (-\,0{\cdot}592)^2 + (-\,1{\cdot}02)^2 + (-\,1{\cdot}18)^2\\ &+ (-\,1{\cdot}02)^2 + (-\,0{\cdot}592)^2 + (0)^2 + (0{\cdot}592)^2\\ &+ (1{\cdot}02)^2 + (1{\cdot}18)^2 + (1{\cdot}02)^2 + (0{\cdot}592)^2]\\ = 0{\cdot}836&(N_s/N_p)I \text{ or } 0{\cdot}68(V_s/V_p)I\end{aligned}$$

The supply loading with a delta-connected primary is then

$$\begin{aligned}\sqrt{3}V_L I_L = \sqrt{3}V_p I_L &= 3(N_p/N_s)V_2 \times 0{\cdot}836(N_s/N_p)I \\ &= \sqrt{3} \times 0{\cdot}816 \times 0{\cdot}86E_0 \times 0{\cdot}836I \\ &= 1{\cdot}01E_0 I\end{aligned}$$

Voltage Fluctuation

It will be noted from Fig. 14–15 that the two six-phase groups combine to produce a fluctuation in the output voltage of twelve times the mains frequency, the amplitude of the fluctuation being

$$\begin{aligned}\sqrt{2}V_s \sin \pi/3 - (\sqrt{2}V_s \sin 5\pi/12 + \sqrt{2}V_s \sin \pi/4)/2 \\ = 0{\cdot}0417V_s \text{ or } 0{\cdot}0355E_0\end{aligned}$$

CHAPTER XV

GRID-CONTROLLED RECTIFIERS, INVERTERS, AND FREQUENCY CHANGERS

An extremely important development in the rectification field is the application of grid control to the mercury-arc rectifier. The principles involved are similar to those of the thyratron and, strictly speaking, a grid-controlled mercury-arc rectifier may be termed a thyratron. Various methods of grid control are employed, the purpose of which is to permit the establishment of the arc at some predetermined moment during the anode voltage cycle by the application of a positive potential to the grid, and to prevent re-establishment of the arc at some other moment by the application of a negative grid potential. The latter potential is termed the blocking potential and the former the liberating potential. In practice, the values of these potentials vary with the design and voltage rating of the rectifier, the blocking potential usually lying between 25 and 250 volts and the liberating potential between 25 and 150 volts. In order to limit the value of the grid current, series resistances are employed. In the case of steel-tank rectifiers the grid current is of the order of 200 to 500 mA, while with glass-bulb rectifiers the current lies between 10 and 50 mA.

Soft-control Systems

Grid-control systems may be classified under two headings, namely soft-control systems and hard or impulse systems. Considering the first-named, we may introduce this by means of the simple arrangement indicated by Fig. 15–1. Referring to this, let it be assumed that we have a positive-control tube the critical grid potential of which is constant as shown. Thus, for the case assumed, the grid-excitation curve is a straight line parallel to and above the abscissa. If now a sinusoidal e.m.f. is impressed on cathode and grid, it is evident that when the instantaneous value of this e.m.f. is equal to the critical grid voltage, the tube will strike and anode conduction will be established. Assuming that Fig. 15–1 represents

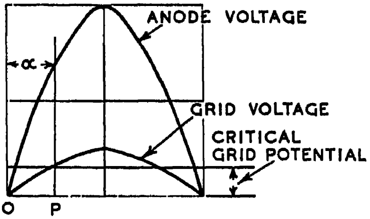


FIG. 15–1

a single-phase rectifier, it is clear that the effect of grid control is to delay striking from O to P. The angle corresponding to the distance OP, measured along the base line, is known as the ignition angle and designated by α. It is evident that for the case considered α may be continuously increased from 0° to 90° and that any attempt to increase α beyond this results in failure to establish an arc.

Before proceeding further, the mean value of the current for the case represented by Fig. 15–1 will be determined, on the assumption that we are dealing with a half-wave rectifier. Assuming a pure resistance load, the value of which is R, the instantaneous value of the d.c. current is

$$\frac{\sqrt{2}V \sin \theta - E_a}{R}$$

where E_a is the arc drop of the rectifier. The mean value of the current is

$$I = \frac{1}{2\pi R}\int_{\alpha}^{\pi - \theta_1} (\sqrt{2}V \sin \theta - E_a) d\theta$$

where $\theta_1 = \text{arc} \sin E_a/\sqrt{2}V$. Integrating,

$$I = \frac{\sqrt{2}V}{2\pi R}\left[\cos \theta_1 + \cos \alpha - \frac{E_a}{\sqrt{2}V}(\pi - 0 - \alpha)\right]$$

which, if $E_a/\sqrt{2}V$ is small, becomes

$$I = \frac{\sqrt{2}V}{2\pi R}(1 + \cos \alpha) \quad . \quad . \quad . \quad . \quad . \quad . \quad (15\text{–}1)$$

The rectified voltage across R is, of course,

$$E = IR = \frac{\sqrt{2}V}{2\pi}(1 + \cos \alpha) \quad . \quad . \quad . \quad . \quad . \quad (15\text{–}2)$$

For the case of full-wave rectification (15–1) and (15–2) respectively become

$$\frac{\sqrt{2}V}{\pi R}(1 + \cos \alpha)$$

and

$$\frac{\sqrt{2}V}{\pi}(1 + \cos \alpha)$$

(15–1) and (15–2) immediately show that the value of the rectified current may be controlled by the value of α, i.e. the point at which the grid-cathode voltage intersects the critical grid potential. This, of course, is true whether a positive or negative-control rectifier is employed and is independent of the method of initiating the arc. The conditions for a negative-control rectifier are indicated by Fig. 15–2, the critical grid potential curve being drawn from Fig. 9–10. It is again evident that the maximum value of α is 90°.

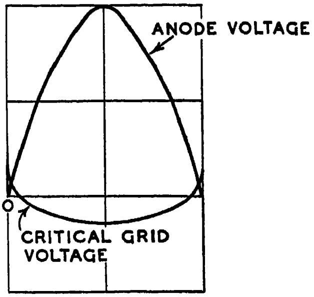


FIG. 15–2

Phase-shift Control

Although a simple sinusoidal grid-cathode method of control has been assumed for the purpose of introducing grid control, this is seldom employed because of associated disadvantages. For example, although (15–1) indicates that I may be continuously varied from a maximum, corresponding to $\alpha = 0$, to zero, with the method of control assumed α may only be varied from 0° to 90°. At 90° the current suddenly fails and it will be appreciated that the current may only be varied from a maximum to one-half this value. This effect is shown by Fig. 15–3.

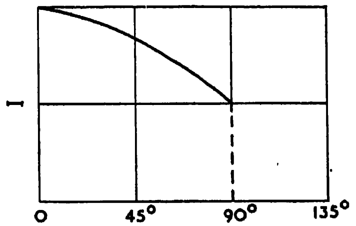


FIG. 15–3

In order to effect continuous variation of I over 0° to 180°, what is known as phase-shift control may be employed. This is achieved by varying the phase angle between the grid and anode potentials. Referring to Fig. 15–4, the grid potential is assumed to be lagging behind the anode potential by θ degrees. At α the grid voltage intersects the critical grid potential curve and conduction occurs. It is evident that if the phase of the grid potential may be varied between 0° and 180°, I may be continuously varied between a maximum and zero for these limits. In the event of the magnitude of the grid voltage being large compared with the critical grid potential, $\alpha \doteqdot \theta$ and the critical grid curve may be assumed to coincide with the zero voltage axis.

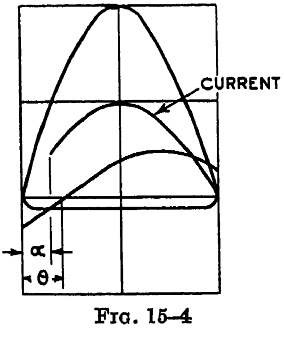


FIG. 15–4

Furthermore, if $E_a/\sqrt{2}V$ is small, then (15–1) gives the relation between I and α for all angles, up to 180°, by which the grid voltage lags behind that of the anode. Consideration of Fig. 15–4 shows that if the grid voltage leads the anode voltage by any angle, conduction occurs almost at the commencement of each cycle and I is unaffected by the phase or magnitude of α. Hence as α passes from 180° to a value slightly in excess of this, I will suddenly rise from zero to a maximum in the manner indicated by Fig. 15–5.

The foregoing considerations of phase-shift control have been considered for a negative-control tube. However, it is apparent that they are equally applicable to positive control tubes.

PHASE-SHIFTING CIRCUITS

In order to employ phase-shift control some form of phase-shifter is necessary. For large polyphase rectifiers a polyphase phase-shifter is generally employed. This virtually consists of an induction motor possessing a wound rotor, the latter being capable of limited rotation. The stator winding is excited from a three-phase supply and hence produces a rotating field, which induces e.m.f.s in the rotor windings. The latter are connected to the rectifier grids, and by altering the relative positions of stator and rotor the phase angle between grid and anode voltages may be varied.

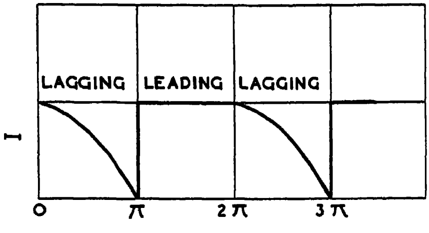


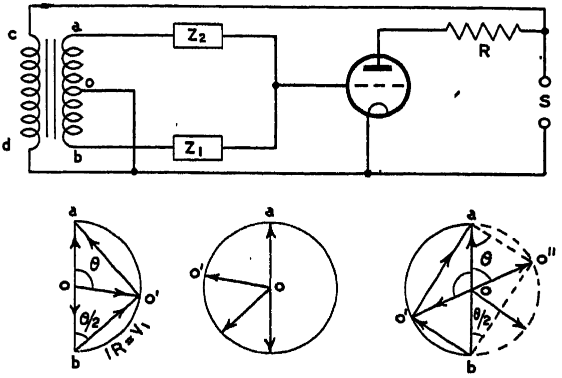
FIG. 15–5

A further method of phase-shifting is indicated by the network of Fig. 15–6 and the accompanying vector diagram. The cathode is connected to the electrical centre of a transformer secondary, the primary of which is excited from the main supply voltage. Assuming that Z_1 and Z_2 are, respectively, a resistance R and an inductive reactance wL, their voltage vectors will be at right-angles and are denoted by V_1 and V_2. The vector sum of these two voltages is, of course, constant and equal to $2V$, where V is one-half of the transformer secondary voltage. The cathode being connected to the transformer centre and the grid to the junction of Z_1Z_2, the grid-cathode voltage vector extremities must fall on the centre of $2V$ and the point of intersection of the vectors V_1V_2. Hence the grid-cathode voltage is constant and equal to V and is unaffected by variation of its phase relation to, say, the vector $2V$. Hence by varying either R or L the phase of the grid voltage may be varied from 0° to 180°. Assuming now that the transformer is appropriately connected, the grid-cathode voltage will lag behind the anode voltage. It will be constant in magnitude and variable in phase as already described. Referring to Fig. 15–6 (*a*), we may note

$$\tan\frac{\theta}{2} = \frac{V_2}{V_1} = \frac{Z_2}{Z_1} = \frac{wL}{R}$$

It follows that if R is the variable, for $\theta = 180°$, $R = 0$, and for $\theta = 0$, $R = \infty$.

If Z_1 and Z_2 are interchanged, this has the effect of reversing the vector OO' and causing the grid voltage to lead on that of the anode, with the result that no control of the output is possible. This is shown by Fig. 15–6 (*b*) for the *potential difference* between OO' now becomes the *vector difference* of Oa and bO', or the vector sum of Oa and bO' reversed. Formerly it was, of course, the vector difference of Oa and $O'a$. However, no particular care need be taken with regard to the positions of Z_1 and Z_2 in setting up the circuit, for should an angle of lead be obtained this can be simply reversed by reversing the connexions of the primary of the transformer. A condenser may, of course, be used in place of an inductance. Replacing Z_2 by a condenser gives the vector diagram (*c*) in which OO' leads Oa. As this gives no control, Z_2 and Z_1 must be interchanged, when OO' will be reversed, giving the vector diagram shown dotted. From this it is apparent that, in contrast to when an



(a) (b) (c)

Fig. 15–6

inductance is employed, an increase in Z_1, i.e. R, causes a decrease in current. Also

$$\tan\frac{\theta}{2}=\frac{Z_1}{Z_2}=\frac{R}{1/wC}=wCR$$

Fig. 15–7 shows the relation of current to angle of lag for the thyratron of Fig. 9–10 when employed with the following components:

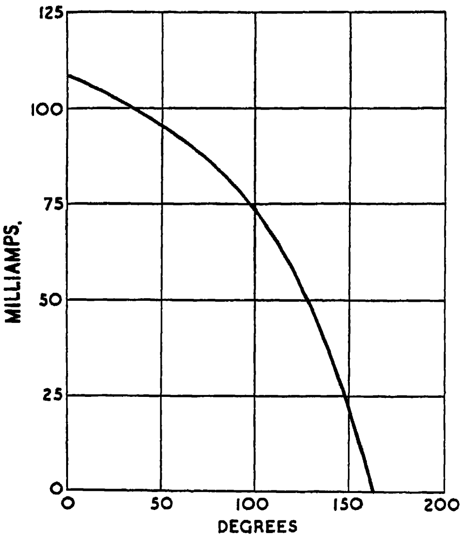


FIG. 15–7

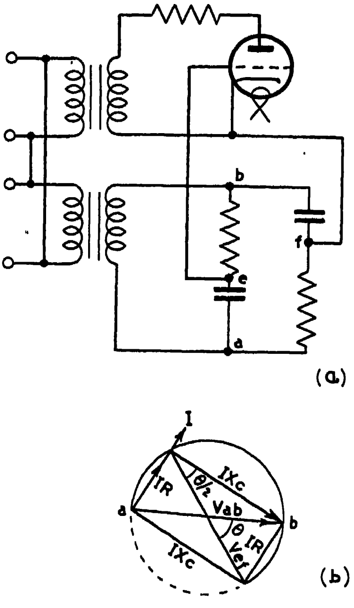


FIG. 15–8

$C = 4$ μF, $R = 1000$ ohms, supply voltage, S, 240 volts, r.m.s.; *ab*, 12 volts, r.m.s. It will be noted that the current falls to zero before $\theta = 180°$. This is due to the arc extinguishing when the supply voltage falls below the arc voltage, i.e. 16 volts.

A further form of phase-shifting circuit is shown by Fig. 15–8 (*a*), and takes the form of a bridge network where the two resistances and condensers are of identical value. Referring to the left-hand branch of the bridge, the vector diagram is shown by the circle diagram at (*b*). The current, of course, leads the transformer voltage vector *ab*, as shown, the latter being the sum of the vectors IR and IX_c, where X_c is the reactance of C. Now, from a consideration of Fig. 15–6 (*c*) the potential of the point *e*, Fig. 15–8, to an imaginary centre tap on the transformer is given by OO', Fig. 15–6 (*c*). Similarly the potential of the point *f* is equal to OO' reversed. Hence the potential difference between *e* and *f* is given by 2 OO', and this is equal in magnitude to *ab*. Hence, for the circuit of Fig. 15–8, the vector diagram is given by (*b*), θ being the angle between grid and anode voltages. Also

$$\tan\frac{\theta}{2} = \frac{R}{X_c} = \omega CR$$

Bias-shift Control

It is, of course, possible to control the moment of arc ignition, in the case of a negative-control tube, by means of a variable d.c. bias voltage. Referring to Fig. 15–9, if E_g is the value of the steady negative bias voltage the point of intersection of this with the critical grid curve gives the moment of ignition. It is evident that control can be effective only from approximately 0° to 90° and that the output current will be varied from a maximum to half-maximum, followed by sudden extinction.

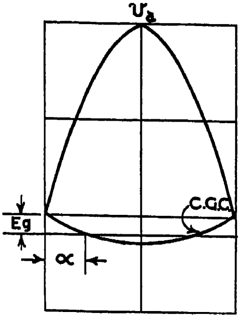


FIG. 15–9

In order to overcome the limitations of the foregoing method of control, what is termed bias-phase or bias-shift control is employed. This consists of combining a d.c. and an a.c. voltage and applying these to the grid. The relative phase of the a.c. component remains constant with regard to the anode voltage, while the magnitude and polarity of the d.c. component are capable of variation. Bias-shift control may be applied to either positive or negative control-tubes and its principle may be appreciated from Figs. 15–10 and 15–11. Referring to Fig. 15–10, the purpose of *R* and *C* is to provide a fixed phase-shift between grid and anode

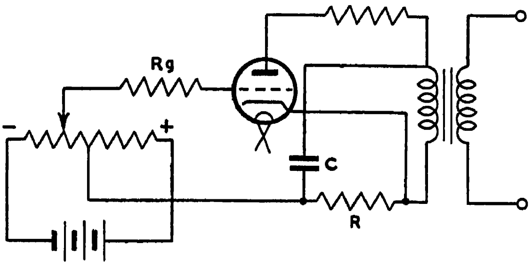


FIG. 15–10

voltages. Thus, the voltage on R is combined with the d.c. voltage from the battery and the combination applied, via the potentiometer, through the grid resistance, R_g, to the grid.

If R and C are such that $R = X_c$, then the a.c. grid voltage will lag 45° behind the plate voltage. Assuming the critical grid curve coincides with the zero line and that $E_g = 0$, the ignition will occur for $\alpha = 45°$ as shown by Fig. 15–11 (a). If d.c. negative bias is now applied, the a.c. voltage, the abscissa of which is now lowered by an amount E_g, will intersect the zero line later in the cycle, as shown at (b), resulting in a shorter conduction period and a reduced output. By progressively increasing the negative bias it is clear that a condition will be reached at which the peak a.c. grid voltage is just equal to the d.c. negative bias. Under this condition the output current is a minimum and any further increase in bias results in failure to establish an arc. For the case shown this occurs for $\alpha = 135°$.

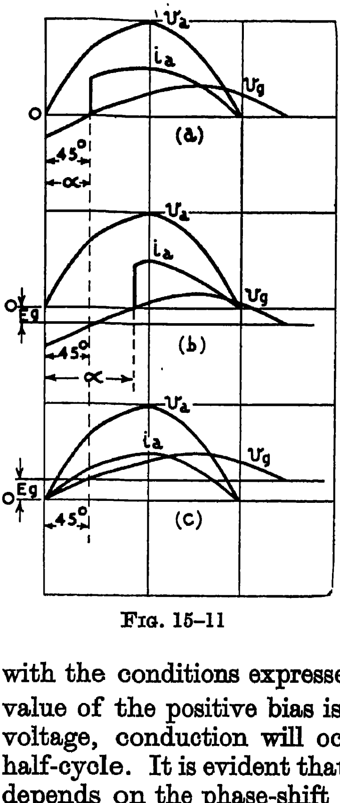


FIG. 15–11

If a positive bias is applied, the a.c. grid voltage will intersect the zero line earlier in the cycle, as shown by (c). This, of course, increases the conduction period and output current compared with the conditions expressed by (a). For the case shown, when the value of the positive bias is equal to $1/\sqrt{2}$ times the peak a.c. grid voltage, conduction will occur at zero and continue over the full half-cycle. It is evident that the range over which control is effective depends on the phase-shift angle. For the case under consideration effective control is obtained over 135°. It is apparent that control over the full half-cycle occurs for $\alpha = 90°$, a condition which, with the phase-shift circuit shown, is practically impossible of attainment.

GRID-CONTROLLED POLYPHASE RECTIFIERS

The foregoing phase-shifting circuits are principally applicable to half-wave and full-wave rectifiers. Phase-shift control applied to polyphase rectifiers, as mentioned on page 492, is shown by Fig. 15–12. The phase-shift transformer is of relatively low capacity, usually less than 0·1 kVA, and control is simply effected by altering the relative positions of stator and rotor. The system readily lends itself to remote control.

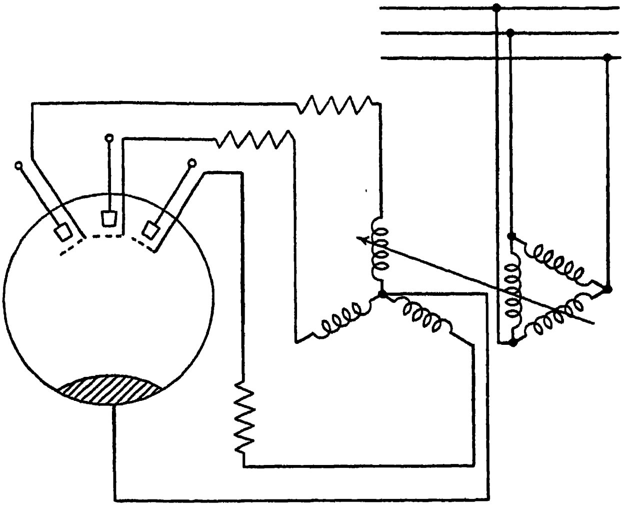


FIG. 15–12

Bias-shift control as applied to polyphase rectifiers is shown by Fig. 15–13. A small potential transformer is employed, the sliding arm of the potentiometer being connected to the neutral point of the secondary winding. The potentiometer can, of course, be readily adapted to remote control. As previously shown, the range over which control may be effected depends on the phase-shift angle.

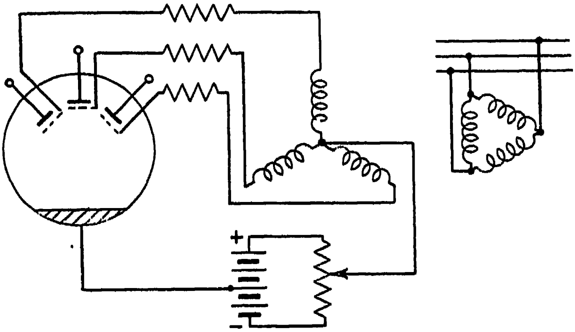


FIG. 15–13

The foregoing methods of controlling polyphase rectifiers are, of course, soft-control methods. Hard, or impulse, systems of control are shown by Figs. 15–17 and 15–18.

RELATION BETWEEN OUTPUT VOLTAGE AND IGNITION ANGLE

We shall now derive an expression for the relation between the d.c. output voltage and the ignition angle α for the case of the polyphase rectifier. Reference to Fig. 15–14 shows the no-load voltage waves of an m-phase rectifier where, in the absence of bias, commutation commences at A and finishes at B. In these circumstances, as we have already seen, the mean d.c. voltage is given by

$$E_0 = \sqrt{2}V_s \times \frac{m}{2\pi}\int_{\frac{m-}{2m}\pi}^{\frac{m+2}{2m}\pi} \sin \theta d\theta$$

$$= \sqrt{2}V_s \cdot \frac{m}{\pi} \sin \frac{\pi}{m}$$

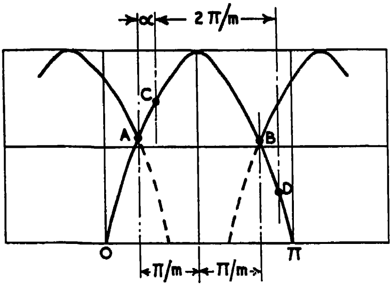


FIG. 15–14

If commutation is retarded by an angle α, then the commutating points are C and D, the interval CD being identical with AB and equal to $2\pi/m$. In this case we have

$$E_{0\alpha} = \sqrt{2}V_s \times \frac{m}{2\pi}\int_{\frac{m-2}{2m}\pi+\alpha}^{\frac{m+2}{2m}\pi+\alpha} \sin \theta d\theta$$

$$= \sqrt{2}V_s \cdot \frac{m}{\pi} \sin \frac{\pi}{m} \cos \alpha$$

$$= E_0 \cos \alpha$$

where $E_{0\alpha}$ is the no-load voltage corresponding to an ignition angle α.

Now if there is no inductance present, the angle α is restricted because the upper integration limit above cannot exceed π for, at this point, the anode-cathode potential difference becomes negative and, in the absence of inductance, the arc is extinguished. Hence in the non-inductive case α may be continually increased until

$$\frac{m+2}{2m}\pi + \alpha = \pi$$

at which time D will reach π. Having reached this point α, as measured from A, Fig. 15–14, may be increased until

$$\frac{m-2}{2m}\pi + \alpha = \pi$$

at which time $E_{0\alpha}$ will become zero. The value of α, say α_D, when D reaches π is given by

$$\frac{m+2}{2m}\pi + \alpha_D = \pi$$

and

$$\alpha_D = \left(\frac{1}{2} - \frac{1}{m}\right)\pi$$

If α is increased beyond this value, the direct current will not pass continuously through the load because anode conduction will cease between the angles α_D and α. Hence if $\alpha > \alpha_D$ the load current consists of portions of sine curves separated by zero-current regions. Assuming $\alpha > \alpha_D$ the mean d.c. output voltage is given by

$$E_{0\alpha} = \sqrt{2}V_s \times \frac{m}{2\pi}\int_{\frac{m-2}{2m}\pi+\alpha}^{\pi} \sin\theta d\theta$$

$$= \sqrt{2}V_s \frac{m}{2\pi}\left[1 - \sin\left(\alpha - \frac{\pi}{m}\right)\right]$$

with a final limit for α of $\pi\left(\frac{1}{2} + \frac{1}{m}\right)$ at which value the d.c. voltage becomes zero.

Generally speaking it is undesirable in practice to allow the current to fall to zero and it is, therefore, usual to include an inductance in the d.c. circuit to prevent this happening. It will be noted that an effect of grid control is to increase distortion of the d.c. output voltage, particularly if the voltage range over which control is exercised is large. Where this is objectionable, voltage variation is sometimes effected by means of a tapped transformer. The advantages of grid-control are, however, so obvious that it is scarcely necessary to detail these. A particularly useful feature in certain applications is that the voltage may be gradually raised from zero to the normal value. In this manner surges may be avoided which might otherwise occur.

CIRCUIT INTERRUPTION BY GRID CONTROL

A valuable feature of the grid-controlled rectifier is the possibility of causing it to act as its own circuit-breaker. Thus, in the event of a short-circuit or back-fire, the automatic application of a negative potential to the grids rapidly causes an interruption of the d.c. supply, thus safeguarding the rectifier installation. This, of course, obviates the necessity of high-speed circuit-breakers on the direct-current side and large capacity oil-filled switches on the alternating-current side. In the case of a rectifier in which grid

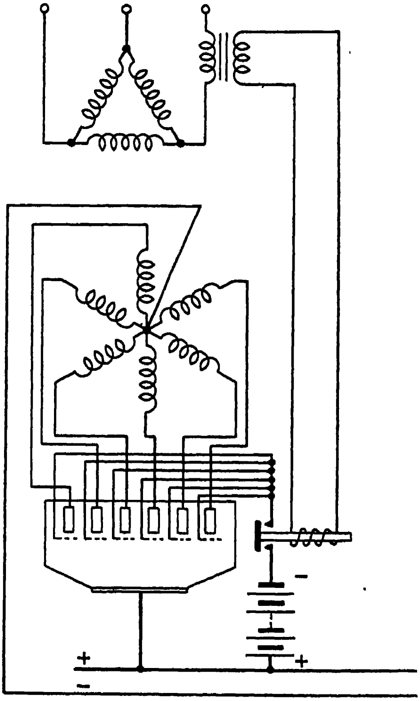


FIG. 15–15

control of the voltage is not normally employed, current interruption is effected by the sudden application of a negative, or blocking, potential to the grids, the latter, while a fault does exist, being negative. A system of this kind is indicated by Fig. 15–15, where a high-speed relay may either be energized by a diverter in the d.c. circuit or a current transformer (as shown) in the a.c. circuit. Should an overload or back-fire occur, the relay is closed, thus applying a blocking potential to the grids. No anode can now pick up the arc, with the result that the current can persist only until the end of the half-cycle at that anode to which it was passing at the commencement of the overload. Thus, in the case of a three-phase rectifier, the maximum period of overload is 150° and in the six-phase case 120°, these periods respectively corresponding to $8\frac{1}{3}$ and $6\frac{2}{3}$ milliseconds.

A system of circuit interruption applicable to rectifiers normally employing grid control for voltage-regulating purposes is shown by Fig. 15–16. In this case an overload has the effect of opening the relay, thus interrupting the supply of positive impulses which, normally, are essential for liberating the grids.

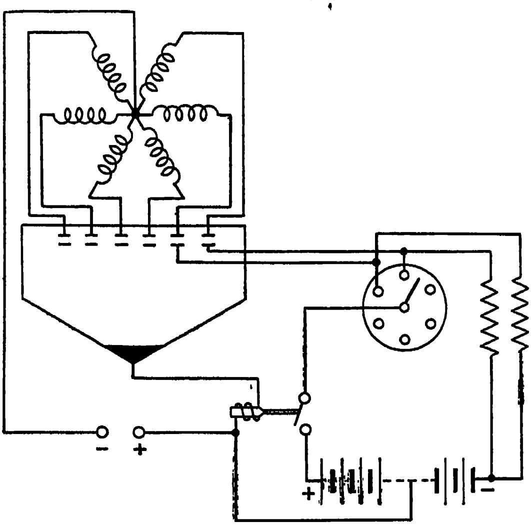


Fig. 15–16

Hard Control Systems

The foregoing methods of grid control, when applied to mercury-vapour rectifiers, suffer from the disadvantage that they do not permit precise determination of the moment of arc ignition. This is evident from Fig. 9–19, for from this we see that the critical grid potential is a function of the mercury-vapour temperature. Hence the critical grid curve depends on the temperature and loading of a rectifier. For example, from Fig. 9–19 it will be noted that an increase in mercury-vapour temperature lowers the critical grid potential, thus causing an earlier intersection of the cathode-grid potential with the critical grid curve. This reduces α, causing an increase in output voltage. From this we may infer that an increase in loading may cause an increase in output voltage. Referring to Fig. 15-1, it is evident that the effect of a change in loading and temperature may be reduced if the obliqueness with which the grid voltage intersects the critical grid curve is increased.

From the foregoing it follows that in order to obtain precise control of the moment of arc ignition, potential impulses must be applied to the grid of such character as to intersect the grid curve practically at right-angles. Reference to Fig. 15-1 will show that in these circumstances variation in the critical grid voltage has an inappreciable effect on the ignition angle. A manner in which this may be effected is indicated by Fig. 15-17, which shows impulse control applied to a half-wave rectifier. The method consists of alternately applying negative and positive grid impulses by means of a synchronously-driven contact-breaker. It will be seen that while the contact is open the grid is negative to the cathode by e_1 volts. On contact closure a current flows in r_1r_2 given by

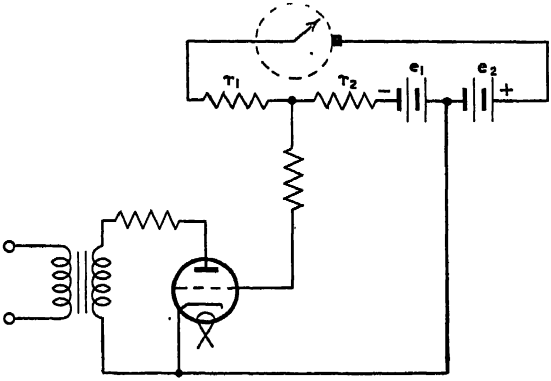


FIG. 15-17

$$i = \frac{e_1 + e_2}{r_1 + r_2} \text{ amperes}$$

and hence the voltage drop on r_2 is

$$\left(\frac{e_1 + e_2}{r_1 + r_2}\right) r_2 \text{ volts}$$

Thus the grid is now positive to the cathode by

$$\left[\frac{e_1 + e_2}{(r_1 + r_2)} r_2 - e_1\right] \text{volts}$$

These voltage changes, of course, occur almost instantaneously and in consequence the moment of ignition is particularly precisely defined. This moment is varied as desired merely by rotating the normally fixed contact with respect to the rotating one.

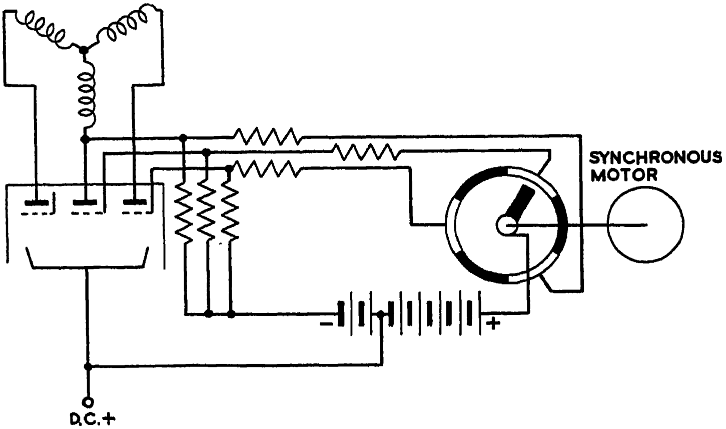


FIG. 15–18

The application of the foregoing scheme of impulse control to a polyphase rectifier is shown by Fig. 15–18 and the conditions obtaining by Fig. 15–19. It will be noted that the positive impulse need be only of short duration, this being arranged by the time of

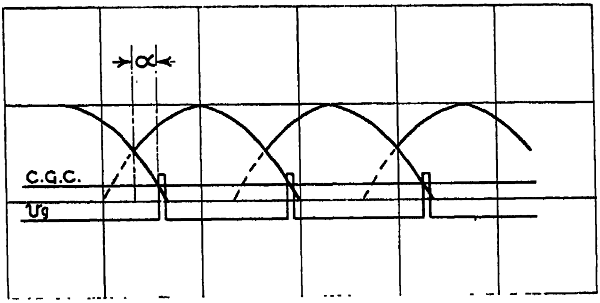


FIG. 15–19

overlap of the distributor contacts. Where only manual control of the output is required, this is effected simply by moving the distributor. If remote or automatic control of the output is desired, the angular position of the distributor is adjusted by means of a small motor operating through a worm gear.

A further method of hard, or impulse, control employs grid-excitation transformers possessing a peaky secondary voltage wave for providing the potential impulses. This is effected by arranging conditions to be such that the transformer cores are saturated. The general arrangement as applied to a three-phase rectifier is shown by Fig. 15–20. From this it will be seen that as many transformers are needed as there are anodes. Of the three windings on each transformer, windings 1 are connected in delta, while windings 2 are star-connected through a variable resistance R_2. Each delta winding carries a sinusoidal magnetizing current which is limited by the resistance R_1. As the voltages on windings 1 and 2 have a phase displacement of 150°, it follows that the star-connected windings carry a magnetizing current which is displaced in phase from that of the delta windings by approximately 150°. The magnitude of this current is controlled by the resistances R_2. Assuming the grid-transformer magnetizing currents lag behind the voltages of the main rectifier transformer by angles θ_1 and θ_2, the conditions may be represented by the vector diagram of Fig. 15–21, where m_2 and m_1 are, respectively, the m.m.f.s of the delta and star-connected windings and m_r the resultant m.m.f. The flux due to the latter

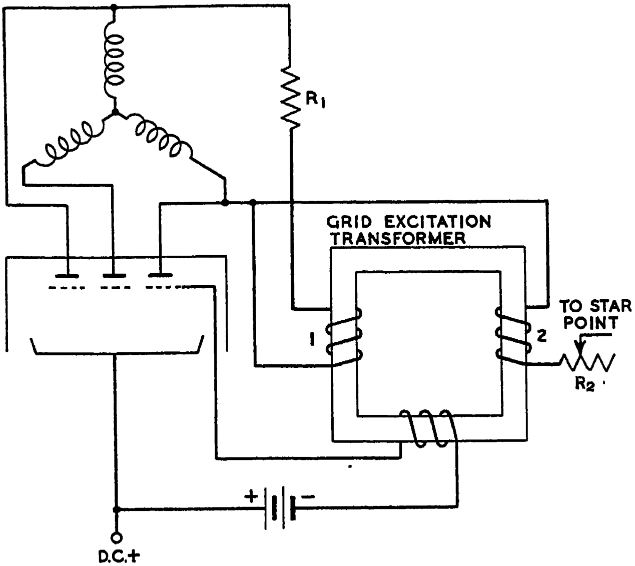


Fig. 15–20

induces an e.m.f., e_3, 90° ahead of itself in winding 3, this e.m.f. being of a peaky waveform due to saturation of the grid-transformer core. Hence by varying the magnitude of m_1 and m_2 (by means of R_1 and R_2) the phase of the e.m.f. in winding 3 may be varied, and, consequently, the angle of ignition.

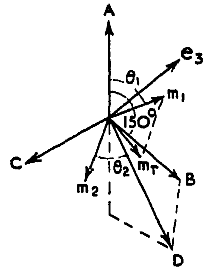


Fig. 15–21

The mode of operation of the foregoing system of control will be clear from Fig. 15–22. The grids are normally held negative by a steady d.c. voltage upon which is superimposed the positive impulses of the grid-excitation transformers. The resultant m.m.f. is shown by the curve m_r, and the flat-topped flux wave by ϕ. It will be noted that the potential impulses occur as m_r and ϕ are passing through their zero values. Whether a negative or positive control rectifier tube is employed (a positive type is shown), it is evident that the potential impulses intersect the critical grid curve sufficiently steeply to render the moment of ignition independent of variations in the critical grid potential. Following a positive impulse from any transformer is a negative impulse from the same transformer 180° later. However, this is immaterial in the operation of the system.

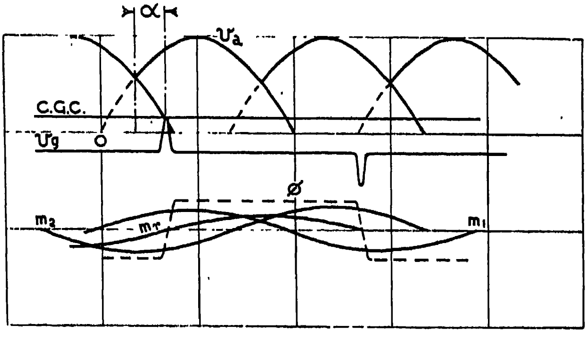


Fig. 15–22

By a modification the foregoing system may be made automatic and a compounding effect produced. Fig. 15–23 shows a manner in which this may be carried out. Instead of the variable resistances, R_2, chokes are employed consisting of coils wound on cores which are saturated by means of two sources. One of these provides a constant d.c. m.m.f., while the other is variable and is derived from a diverter resistance in the d.c. load circuit of the rectifier. As a result, the impedances of the chokes vary with varying load conditions, thus varying the ampere-turns of the grid-excitation transformers. Hence the potential impulses applied to the grids may be varied in phase and any degree or form of compounding produced.

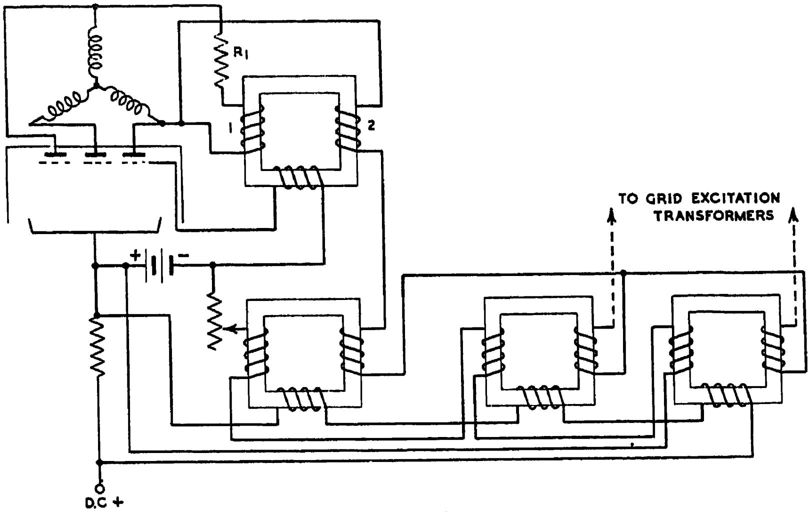


FIG. 15–23

It will be appreciated that such a method is entirely automatic and static, being extremely suitable for traction rectifiers.

The Mercury-arc Inverter

In addition to its normal function of converting alternating to direct current, a rectifier may be caused to effect the reverse process, i.e. convert direct to alternating current. When this occurs the rectifier is termed an inverter, is said to invert, the process being known as inversion. In some respects inversion differs considerably from rectification. For example, with the latter a d.c. voltage is produced from the a.c. source of supply by rectification of the supply voltage. The inverter is unable to produce an a.c. voltage, but only an a.c. current. Hence inversion necessitates an a.c. voltage source, which, however, need only be of small capacity. In addition to providing the necessary voltage, the a.c. supply also provides the

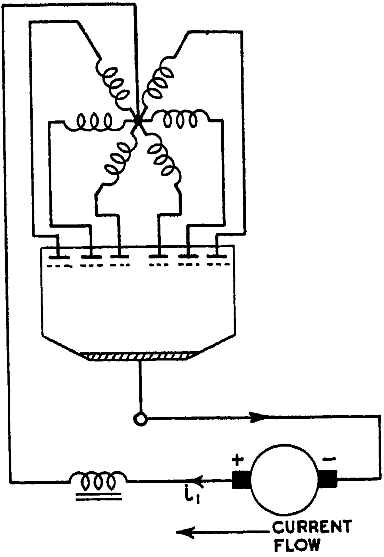


Fig. 15–24(a)

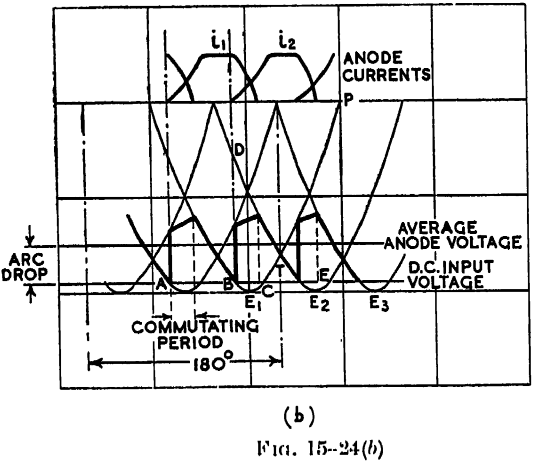


Fig. 15–24(b)

wattless kilovolt-amperes of the arc transformer and fixes the frequency of the inverted current. A further function of the a.c. supply voltage is to extinguish the arc at the correct instants as required by commutation from anode to anode.

In order to present the method by which inversion is effected we shall consider Fig. 15–24 (*a*) and (*b*). From (*a*) it will be noted that the d.c. source has been so connected to a rectifier that the transformer neutral is positive and the cathode negative. This, of course, is necessary in order that the d.c. source may pass current from anode to cathode (conventionally speaking) in the customary way, the reverse direction being, of course, non-conducting. The grids are given a negative potential, with positive impulses occurring at an ignition angle which is so advanced as to cause commutation to occur at some point such as A on the *negative* half-cycle of the anode voltage wave as shown. When ignition occurs, it will be appreciated that due to the magnitude of the d.c. voltage being greater than that of the arc drop and the opposing negative half-cycle, a current will be forced against the latter through the appropriate anode and transformer secondary winding. It will be seen that the negative anode voltage constitutes a back e.m.f. to the current supplied from the d.c. source, and hence the product of the values of this voltage and current is a measure of the power being transferred from the d.c. to the a.c. side of the system. Considering the conditions from the point A, the current i_1 will rise to a maximum at a finite rate due to the presence of the choke. It will then decline as the magnitude of the transformer phase voltage increases towards its maximum. At some point B this voltage will be equal in magnitude to that of the d.c. source, with the result that extinction of the arc will tend to occur. The presence of the choke will, however, cause the current to persist until some point such as C is reached. If at the point B a positive impulse is applied to the succeeding phase, this will commence conducting at D and continue to do so until the point E is reached. Once the arc is extinguished at a point such as C, it remains extinguished because the grid has now a negative bias and restriking cannot occur until a positive impulse is again applied. It will now be appreciated that providing the magnitude of the d.c. voltage is suitable, and positive voltage impulses are applied to the grids at the appropriate moments, commutation will automatically occur, the arc transferring itself from anode to anode in sequence in a manner similar to that in a rectifier.

From the foregoing considerations it is evident that grid control is essential to the process of inversion for, by means of grids, the successive instants of commutation are determined and re-ignition is prevented at unsuitable points in the phase-voltage cycle. Thus, although the current at any anode is naturally terminated when its phase voltage becomes equal to that of the d.c. source, current flow to that anode is forced by means of cyclical positive grid impulses. We see, therefore, that in order to make a rectifier invert, the potential of the d.c. source which it may have been previously supplying must be reversed and the ignition angle, α, retarded. It is also evident that the point to which the ignition is retarded depends on the magnitude of the d.c. voltage. The higher this voltage, the greater is the retardation for a given d.c. current. It is further evident that there is an upper limit for the d.c. supply voltage which generally should be somewhat lower than the maximum value of the transformer secondary voltage wave. In practice, for a given d.c. voltage, the value of α may be initially larger than that actually desired and then gradually reduced until the desired d.c. current is flowing.

The amount which α may be retarded must be less than 180° from the original zero. This is because any one anode must take over the arc while its potential is positive and greater than that of the preceding anode. For example, if the ignition impulse should occur at 180°, i.e. at T, Fig. 15–24, with the d.c. voltage in excess of the peak value of the transformer secondary voltage, the anode preceding that receiving the impulse will still be carrying current. Slightly beyond this point the potential of the anode due to pick up the arc is less than that due to transfer it, with the result that the latter continues to carry the current. The back e.m.f. now continually decreases with the current consequently increasing, and it is evident that the latter may reach a dangerously high value, particularly if the point P should be approached. It is evident that precise grid control is essential with inverters to avoid such possibilities and this implies impulse control.

It is of interest to note that the form of the phase voltages during inversion is that shown by the heavy lines of Fig. 15–24 (*b*). Thus at A, when commutation starts to occur, the phase voltage of the anode relinquishing the arc drops sharply as shown. This anode and that picking up the arc then have a common voltage during the commutation period, at the end of which the phase voltage of the former anode rapidly rises to the appropriate point on the sine wave, while the latter increases towards the point B in the manner indicated. The average phase voltage is shown by a line drawn through the phase voltages, this voltage also being equal to the d.c. input voltage minus the arc drop.

Applications of the Mercury-arc Inverter

Apart from the obvious property of converting direct to alternating current, the inverter has several important applications. One of these is speed and direction control of direct-current motors. The basis of the method of effecting this is shown by Fig. 15–25, in which two identical six-phase rectifiers are cross-connected through the motor undergoing speed and direction control. Referring to Fig. 15–25 and assuming *A* to be rectifying, this rectifier is provided with impulse control, thus permitting any voltage from zero to a maximum to be applied to the motor armature, the motor field being excited from a separate source of supply. *B* is also provided with grid control, the distributor rotors of both *A* and *B* being driven by a common synchronous motor, the distributor of *B* being phase-displaced in the opposite sense to *A* so that if α is the ignition angle of *A*, $(180° - \alpha)$ is approximately that of *B*. Assume the motor is being driven from *A* while *B* has a negative bias only on its grids. Then any motor speed from a maximum to standstill may be obtained by varying the ignition angle of *A*. To slow the motor down, all that is necessary is to suppress the positive impulses on the grids of *A* and apply them to *B*. The latter will now invert, and, providing α is continually decreased as the motor speed falls, regeneration may be obtained practically down to standstill. If α is sufficiently reduced, the motor will again speed up, but in the opposite direction, *B* now acting as a rectifier, while *A* may, if desired, be caused to act as an inverter.

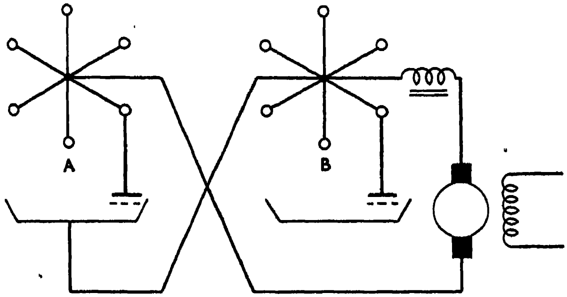


FIG. 15–25

REGENERATION

It is evident from foregoing remarks that the inverter may be used for regenerative purposes. The field to which regeneration is particularly applicable is, of course, traction motors. As the polarity of the traction system must remain fixed, the connexions of the motor cannot be changed and thus the connexions of the rectifier must be reversed. In particular the following sequence of operations must occur if regenerative working is desired consequent on a given percentage rise in d.c. busbar voltage—

(*a*) Disconnexion of the rectifier from the d.c. busbars.

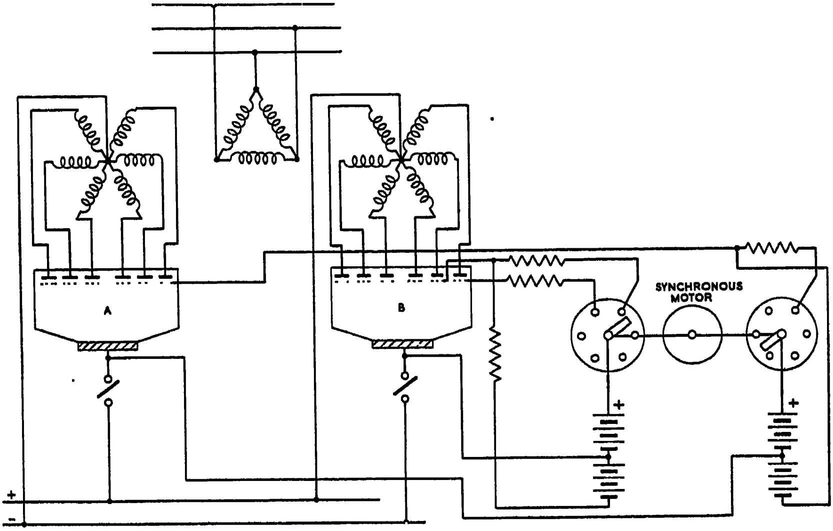


FIG. 15–26

(*b*) Reversal of the rectifier connexions relative to the track system.

(*c*) Adjustment of the grid-control gear, making this appropriate to inverting.

(*d*) Reconnexion of the rectifier to the d.c. busbars.

In order to effect the foregoing, a relay must be employed which will initiate the sequence in either direction as required.

A disadvantage of the foregoing system of regeneration is the difficulty of satisfactorily effecting the above sequence of operations. Because of this, a more usual method of obtaining regeneration is the employment of two units, one acting as a rectifier and the other as an inverter. The scheme is shown by Fig. 15–26, where *A* and *B* represent the two units which are operated from a common transformer primary. The grid distributor rotors are driven by a common synchronous motor, the distributors being arranged in opposite phase. The adjustment of the grid circuits is such that the inverter comes into operation when the d.c. busbar voltage exceeds a predetermined value. Thus, under normal conditions, the positive impulses to the inverter grids would be arranged to occur after the value of the negative-phase voltage was in excess of the d.c. voltage. In these circumstances no driving e.m.f. would be available and inversion would not occur. When the d.c. busbar voltage exceeds a certain figure the positive grid impulses occur before the negative-phase voltage is equal to the d.c. voltage, with the result that regeneration takes place. The amount of energy returned to the supply system depends on the amount by which the d.c. voltage exceeds the input voltage of the inverter. It will be appreciated that the system is entirely static in operation and in this respect is superior to that employing a single unit with relays and reversing switches.

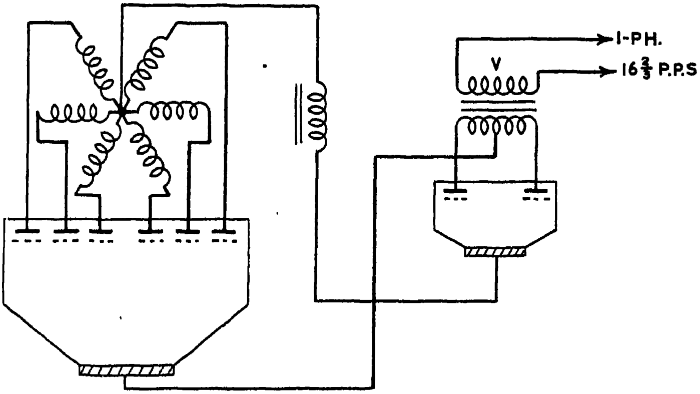


FIG. 15–27

STATIC FREQUENCY CHANGING

An important property of the grid-controlled mercury-arc rectifier is its capability of connecting two supplies of different frequency; for a rectifier reduces the incoming frequency to zero while an inverter raises it to the outgoing frequency. In particular a service for which a rectifier and inverter have proved valuable is the conversion of alternating current from three-phase 50 cycles per second to single-phase 16⅔ cycles per second for traction purposes. One method of effecting this is shown by Fig. 15–27, in which a polyphase rectifier is connected via a large smoothing reactor to a single-phase inverter. The grid-control system of the latter is such as to cause current from the rectifier to flow alternately in each phase of the inverter transformer primary winding. The effect of the reactor is to tend to maintain the d.c. input current at a constant value, thus keeping harmonics in the three-phase supply to a minimum. The constancy of the d.c. current results in the power component of the single-phase supply being of a rectangular waveform. As an inverter cannot generate an a.c. voltage, a synchronous machine is necessary for this purpose and also to fix the frequency of the single-phase system and supply the reactive and harmonic power.

The transfer of the current from one anode to the other in the inverter is, of course, effected in the manner previously described. However, as commutation can occur only while the voltage of the idle anode is higher than that of the working anode, the positive impulse must be applied to the grid of the former just before its voltage passes through a zero value. The ignition angle, α, is indicated by Fig. 15–28 and will be seen to be composed of two parts, u and v. u is the angle of overlap and v is a further period necessary for the de-ionization of the anode relinquishing the current and the consequent regain of grid control. u and v are functions of the load and hence under all load conditions we must have $\alpha > (u + v)$. Assuming the single-phase voltage wave to be approximately sinusoidal, this will result in a sinusoidal load current which, generally, will be inductively reactive. The load current will then consist of two components, the power component I_p and the reactive component I_r. As the latter cannot be supplied from the rectifier it must be furnished by the synchronous machine, as must also be the magnetizing current of the inverter transformer. Denoting the latter by I_{rt}, the total reactive current is $(I_r + I_{rt})$. As the output current from the inverter is assumed sinusoidal and the input current rectangular, it follows that the synchronous machine must also generate the necessary harmonics in order that a sinusoidal output may be obtained. Thus, the total current to be delivered by the synchronous machine is $I = \sqrt{(I_r + I_{rt})^2 + I_h^2}$ where I_h is the harmonic current.

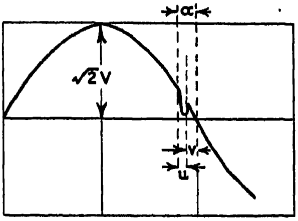


FIG. 15–28

In order to obtain an idea of the conditions prevailing, Fig. 15–29 may be consulted. In this the commutation period has been ignored, the load assumed inductive, and the input current from the rectifier taken to be rectangular in form. The rectifier current is, of course, approximately in phase with the inverter voltage, whereas the load current is lagging behind this as shown. The synchronous machine current is given by the difference of the load and output currents and is shown by curve (*b*).

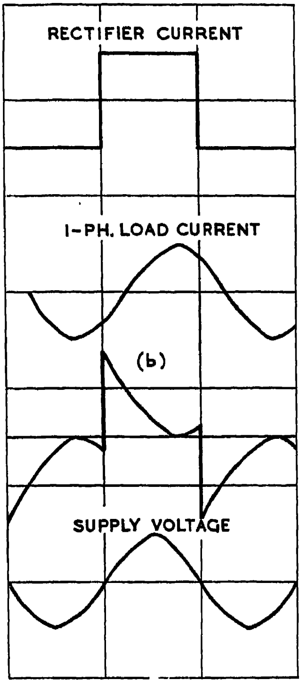


FIG. 15–29

A disadvantage of the foregoing method of frequency conversion is the necessity for a synchronous machine for frequency-fixing and the supply of reactive kilovolt-amperes. The method of frequency conversion to be now described is free from such disadvantages, as it both fixes the frequency and supplies the necessary reactive power. Fundamentally it merely consists of periodically reversing a d.c. potential, a process which, incidentally, in the form of what is known as the vibrator-converter, has been employed for small radio supplies. The method of effecting the reversals may be appreciated by Fig. 15–30. It will be seen to consist of two six-phase grid-controlled rectifiers the secondary windings of the transformers of which are both supplied by a common primary winding. The output terminals of the rectifiers are connected as shown and a single-phase supply of lower frequency is obtained from the two cathodes. The grid control of the rectifiers is such that each comes into operation alternately. For example, the left-hand rectifier operates for one-and-a-half cycles, producing the positive half-wave of the lower frequency single-phase supply, while in a similar manner the negative half-wave is produced by the right-hand rectifier. Hence in the case shown the single-phase frequency is 16⅔ cycles per second.

A disadvantage associated with this method of frequency changing is, of course, the nature of the resulting waveform, which is roughly trapezoidal. The effect of an inductive load is to reduce the harmonics in the waveform, with the result that the load current wave is generally more sinusoidal than that of the voltage. In the case of a non-inductive load the current and voltage waves have identifical shapes and the rectifier secondary-phase currents

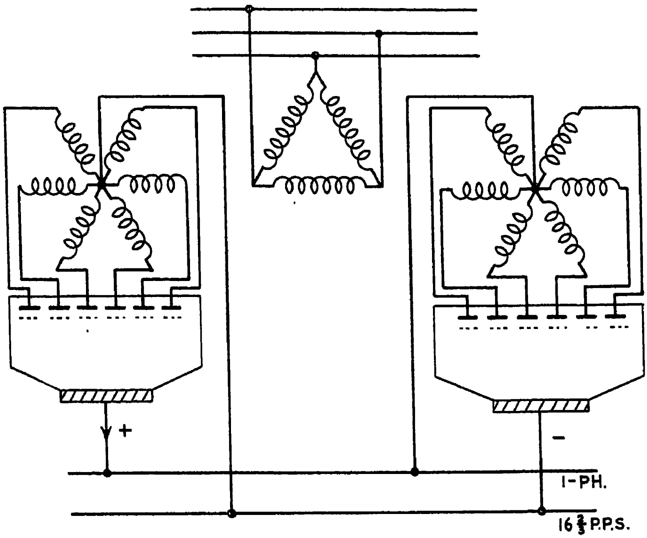


Fig. 15–30 (1)

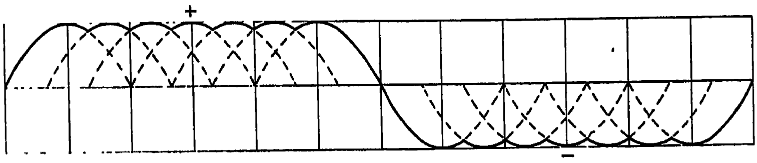


Fig. 15–30 (2)

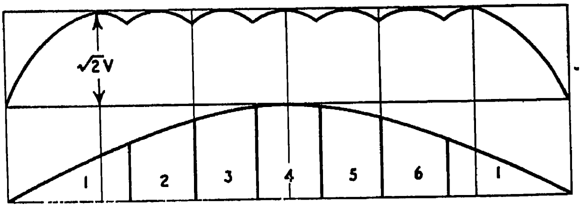


Fig. 15–31

tend to be of rectangular waveform. The voltage and current relationships are shown by Fig. 15–31, where a sinusoidal single-phase load current is assumed. The r.m.s. value of the output voltage is

$$V_1 = \sqrt{\frac{1}{3\pi}\left[2 \times 2V^2\int_0^{2\pi/3}\sin^2 \theta d\theta + 5 \times 2V^2\int_{\pi/3}^{2\pi/3}\sin^2 \theta d\theta\right]}$$

$$= \sqrt{\frac{V^2}{3\pi}\left[4\left(\frac{\pi}{3}+\frac{\sqrt{3}}{8}\right)+\frac{10\pi}{6}\right]}$$

$$= \sqrt{1{\cdot}1V^2} = 1{\cdot}05V$$

Thus $V = 0{\cdot}95V_1$ and the r.m.s. value of the single-phase voltage is 0·75 times the peak value. The instantaneous value of the output current is, of course, equal to the instantaneous value of the various secondary currents. If I_{p1}, I_{p2}, and I_{p3} are, respectively, the r.m.s. values of the primary-phase currents then

$$I^2 = \frac{N_1^2}{N_2^2}(I_{p1}^2 + I_{p2}^2 + I_{p3}^2)$$

where I is the output current. If, as a first approximation, the primary currents are assumed to be equal, then

$$I = \sqrt{3}\frac{N_1}{N_2}I_p \text{ and } I_p = \frac{1}{\sqrt{3}}\frac{N_2}{N_1}I$$

If the primary is delta-connected then $I_L = \sqrt{2}I_p$ and the supply loading is

$$\sqrt{3}E_LI_L = \sqrt{6}E_LI_p$$

$$= \sqrt{6}E_L\frac{1}{\sqrt{3}}\frac{N_2}{N_1}I = \sqrt{6}\frac{N_1}{N_2}V \cdot \frac{1}{\sqrt{3}}\frac{N_2}{N_1}I$$

$$= \sqrt{6}\,0{\cdot}95V_1\frac{1}{\sqrt{3}}I = 0{\cdot}95\sqrt{2}V_1I$$

but V_1I is the single-phase output and, therefore, the distortion factor of the frequency converter is

$$\frac{V_1I}{0{\cdot}95\sqrt{2}V_1I} = 0{\cdot}75$$

which is considerably lower than for a normal six-phase rectifier.

REACTIVE POWER

With the foregoing type of frequency converter when reactive power is drawn from the low-frequency side, it must be supplied at higher frequency on the input side of the converter. Now reactive power is a function of frequency, and if the relation between flux and magnetizing current is linear, then the reactive power varies inversely as the frequency. Hence if I_{Lr} and I_{La} are, respectively, the reactive and active components of the line current, then

$$\tan\phi_h = I_{Lr}/I_{La} \quad . \quad . \quad . \quad (15\text{–}3)$$

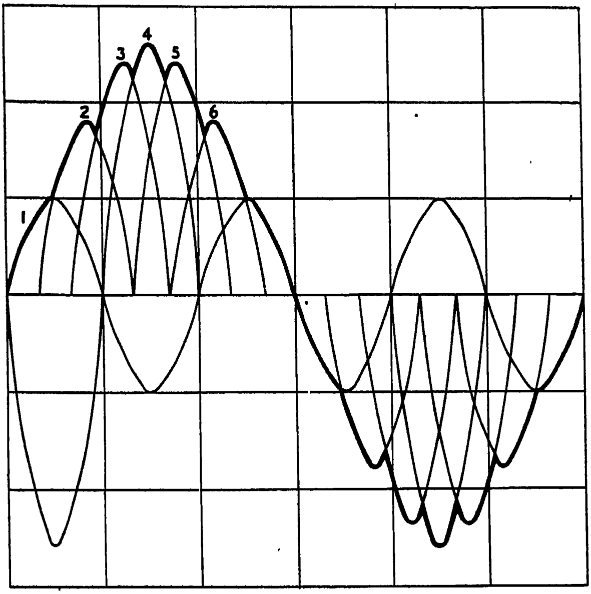


FIG. 15–32

where $\cos\phi_h$ is the displacement factor on the high-frequency side of the converter. On the low-frequency side the ratio of reactive to active power will be three times as great as that expressed by (15–3), and thus

$$\tan\phi_h = \tfrac{1}{3}\tan\phi_l \quad . \quad . \quad . \quad (15\text{–}4)$$

where $\cos\phi_l$ is the displacement factor on the output side. As $\cos^2 x = 1/(1 + \tan^2 x)$, we may write

$$\cos\phi_h = 1/\sqrt{1 + \tan^2\phi_h} = 1/\sqrt{1 + \tfrac{1}{9}\tan^2\phi_l}$$

and the power factor of the converter being the product of the displacement and distortion factors is then

$$0{\cdot}75/\sqrt{1 + \tfrac{1}{9}\tan^2\phi_l}$$

THE ENVELOPE FREQUENCY CONVERTER

The somewhat poor waveform of the previous converter led to a system in which the transformer secondary windings are graded with the object of obtaining a closer approach to a sinusoidal wave. The effect of this will be clear from Fig. 15–32, where the amplitude of each phase voltage on the input side varies as sin θ where θ is reckoned in terms of the lower-frequency cycle. Thus phase 1 has a voltage of 38·27 per cent, phases 2 and 6 have voltages 70·7 per cent, and phases 3 and 5 have voltages 92·39 per cent of that of phase 4. It will be noted that the resulting single-phase wave forms an envelope of the various phases on the secondary side which is, of course, responsible for the descriptive term applied to this form of converter.

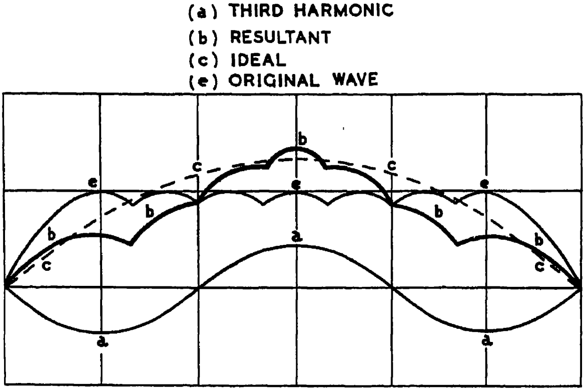


FIG. 15–33

As with the previous converter the frequency ratio is 3 : 1 and

$$\cos \phi_h = 1/\sqrt{1 + \tfrac{1}{9} \tan^2 \phi_l}$$

A further method of improving the waveform of Fig. 15–30 consists of injecting a 33⅓ per cent negative third harmonic into the output voltage. The frequency of this harmonic is, of course, equal to that of the higher-frequency side of the convertor. The general effect of this injection may be appreciated from Fig. 15–33, where it will be noted that a considerable improvement in waveform results. The third harmonic is derived from an induction regulator connected to the three-phase supply, the regulator possessing a single-phase secondary winding. The regulator rotor is adjustable and thus the relative phase position of the injected voltage may be varied and with it the resultant output voltage.

CHAPTER XVI

ELECTRONIC MEASURING INSTRUMENTS

THE limitations of many measuring instruments working on electrodynamic principles have been largely overcome by the employment of instruments operating on electronic principles. A typical example of this is the voltmeter which, in its purely electrodynamic form, has a somewhat high power consumption and, when used on alternating current, possesses waveform and frequency errors. By the employment of magnet materials such as Alcomax, the moving-coil voltmeter may be provided with a resistance as high as 10,000 ohms per volt. Large as this may seem, it is possible to construct electronic voltmeters with input resistances of several megohms with a full-scale deflexion of only a few volts. Furthermore, the latter instruments will function on either a.c. or d.c. supplies and are almost independent of frequency errors.

The number of electronic measuring instruments is now very large and hence only a few basic types can be considered here. They may, however, serve to show the possibilities of this class of instrument. The majority of such instruments usually employ some type of valve, although not necessarily for its rectifying properties.

Valve Voltmeters

A diode-valve voltmeter intended for use on high-frequency supplies is shown by Fig. 16–1. It consists of a high-vacuum diode, a sensitive d.c. galvanometer, and a high resistance *R*, connected in series. If the slope resistance of the valve is small compared with *R*, then the galvanometer reading will be proportional to the mean value of the voltage undergoing measurement. As the same mean value can have different r.m.s. values, i.e. different form factors, it is clear that the instrument will possess waveform errors if calibrated on a waveform different from that under working conditions. The instrument is usually calibrated on a sinusoidal wave of low frequency and is practically independent of frequency errors.

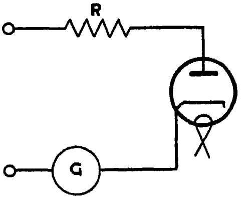


FIG. 16–1

Another form of voltmeter employing a high-vacuum triode is shown by Fig. 16–2. The valve is biased in the manner shown in order to bring the quiescent point down to the lower bend curvature of the anode-current/grid-voltage characteristic. Assuming that this characteristic can be represented by

$$i = av^2 + bv + c$$

then, if a sinusoidal e.m.f. $V_m \sin pt$ is applied between grid and cathode,

$$i = aV_m^2 \sin^2 pt + bV_m \sin pt + c \quad . \quad . \quad (16\text{–}1)$$

where i is the anode current. The mean value of (16–1) over a cycle is

$$\frac{1}{t}\int_0^t (aV_m^2 \sin^2 pt + bV_m \sin pt + c)dt = \frac{aV_m^2}{2} + c$$

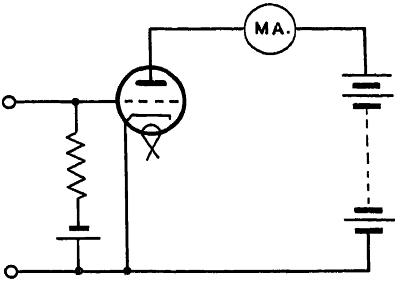


FIG. 16–2

and as c is the anode current when $v = 0$, the increase in the mean value of the anode current due to $V_m \sin pt$ is $aV_m^2/2$. Thus a milliammeter in the anode circuit may be calibrated in volts and will give a deflexion proportional to the square of the applied voltage. The input impedance of the voltage is, of course, equal to the value of R and may be of the order of 10 megohms.

Fig. 16–3 shows a valve voltmeter in the form of a two-valve amplifier, this type of instrument being particularly useful for the measurement of very small a.c. voltages. If μ_1 is the amplification factor of the first valve, then the input voltage to the second valve is

$$V_1\mu_1 \frac{R_1}{R_1 + \rho_1}$$

where V_1 is the input voltage to the first valve and ρ_1 the slope resistance of the latter. Providing the reactance of C is small, the voltage developed across V is

$$V_1\mu_1\mu_2\left(\frac{R_1}{R_1 + \rho_1}\right)\left[\frac{RR_2}{RR_2 + (R + R_2)\rho_2}\right]$$

where μ_2 is the amplification factor of the second valve and R is the voltmeter resistance. If the valves are worked on the linear portions of their characteristics, the voltage across V will be a faithful reproduction of the input voltage. The extent to which the voltmeter readings will be free from waveform and frequency errors will then depend on the voltmeter characteristics only.

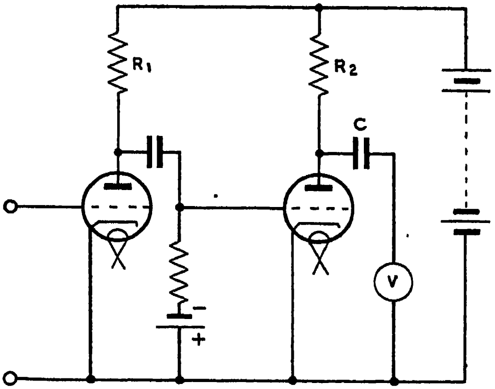


FIG. 16–3

Peak Voltmeters

An electronic voltmeter for measuring peak voltages is shown by Fig. 16–4. This consists of an electrostatic instrument, a diode valve, and a condenser connected as shown. On connecting the voltmeter to an alternating supply, rectification will occur and the condenser will gradually build up to the maximum value of the supply voltage. This maximum will be indicated by the electrostatic voltmeter. Because of leakage the insulation resistance of the entire apparatus must be high and to ensure this it is sometimes necessary to enclose the voltmeter in a heated box which contains some moisture-absorbing substance, such as calcium chloride.

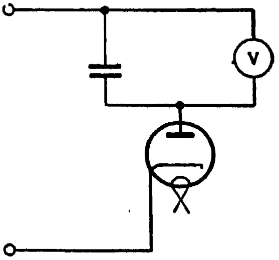


FIG. 16–4

THE SLIDE-BACK VOLTMETER

The principle of this form of voltmeter is shown by Fig. 16–5, where V and G are, respectively, a d.c. voltmeter and a galvanometer. By suitable adjustment of the potentiometer, P, a negative potential may be produced which will just equal the peak value of the positive half-wave of the voltage undergoing measurement. The value of the latter is then given by the reading of V. The method is clearly independent of frequency and waveform and only puts an inappreciable loading on the supply.

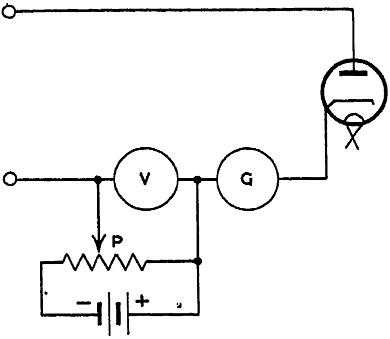


FIG. 16–5

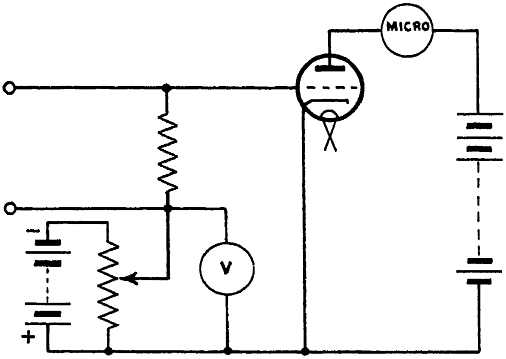


FIG. 16–6

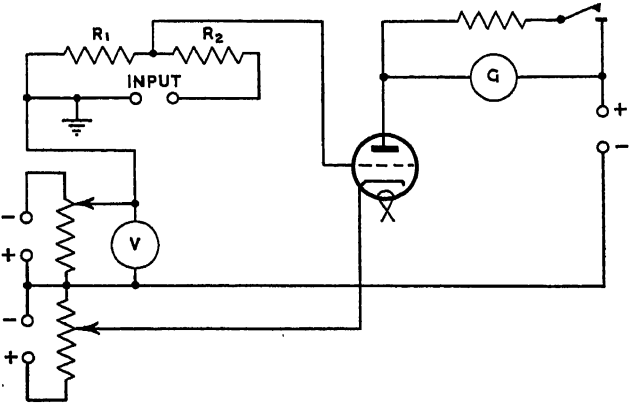


FIG. 16–7

Another example of the slide-back method using a triode is shown by Fig. 16–6. In this case, before applying the voltage to be measured, the valve is biased until the microammeter reads zero. Let the value of the bias voltage to produce this effect be V_1. The voltage to be measured is now applied when the effect of each positive half-cycle will be to deflect the microammeter. The bias is again adjusted for zero deflexion and, if V_2 is the value of this, the peak value of the voltage is $(V_2 - V_1)$ volts. This form of voltmeter has been employed by the writer for measuring magneto voltages up to 20,000 volts, employing the circuit shown by Fig. 16–7, where R_1R_2 is a 20-megohm potentiometer, 20 megohms thus constituting the input impedance of the voltmeter. The complete instrument is shown by Figs. 16–8A and 16–8B (interior view).

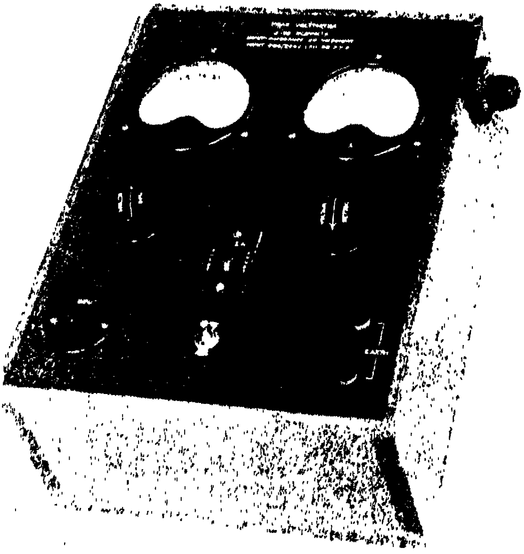


FIG. 16–8A

A further method of peak-voltage measurement is indicated by Fig. 16–9. In order to carry out measurements the switches S_1 and S_2 are both moved to the left, the milliammeter being adjusted to give full-scale deflexion before the voltage to be measured is applied. On application of this, rectification will take place by means of the diode, and a condenser, say, C_1, will be charged to the

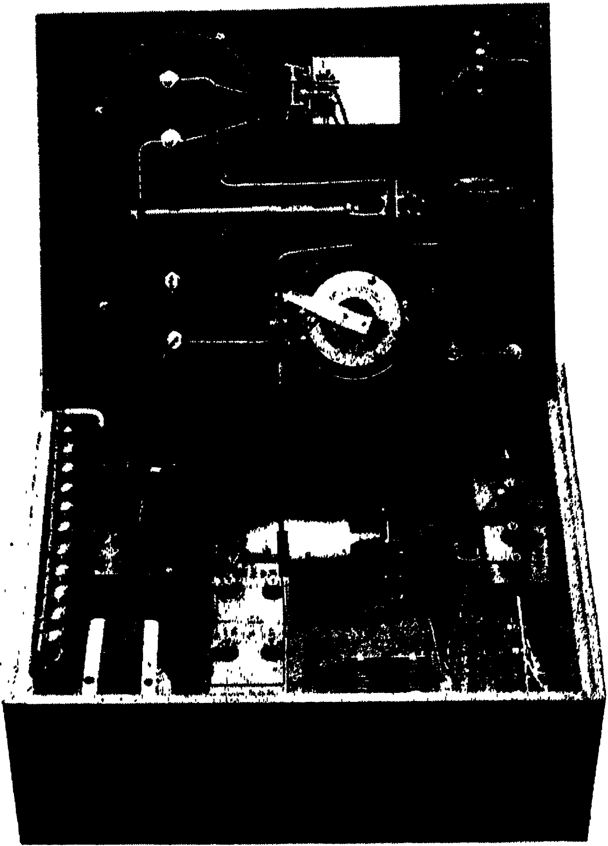


Fig. 16–8b

crest value of the voltage. S_2 is now switched to the right, when C_1 will be placed in parallel with the grid condenser and also the grid and cathode of the triode, thus giving the latter a negative bias and reducing the anode current.

In order to calibrate the instrument, S_1 is switched to the right and a series of known voltages applied to the terminals V. S_2 is then operated as before and the relationship between the milliammeter readings and voltages obtained. A curve may be drawn showing this or the milliammeter may be directly calibrated in volts. To extend the range of the instrument either the transformer ratio may be varied or several condensers used as shown.

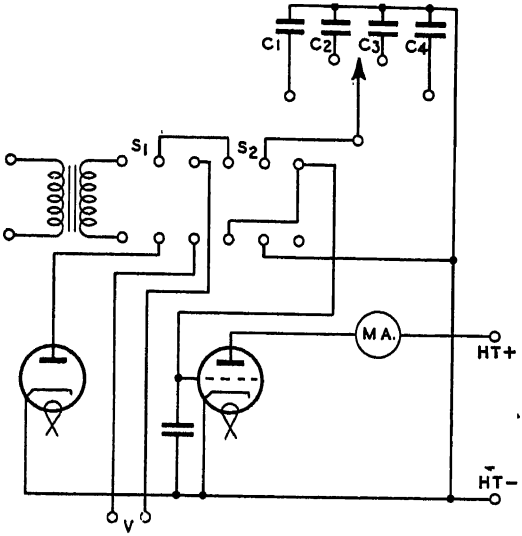


Fig. 16–9

PEAK VOLTAGE MEASUREMENTS WITH THE THYRATRON

For voltage measurements by this method the triode in Fig. 16–6 is replaced by a thyratron. To operate the instrument the bias, V, is set to a value in excess of the peak voltage to be measured, the latter then applied, and the bias then slowly reduced. If V_p is the peak value of the voltage, V_a the anode voltage, and C the control ratio, conduction through the thyratron will occur when $V = (V_p + V_a/C)$. Hence $V_p = V - V_a/C$. The moment that conduction occurs may be indicated by a lamp placed in the anode circuit, this also acting as a current limitor.

THE RYALL CREST VOLTMETER

A method of measuring peak voltages with a neon lamp is shown by Fig. 16–10. Two condensers are employed, one of which is variable. The ratio of the voltages across C_1 and C_2 is inversely proportional to the ratio of their capacitances. Hence if V_s is the striking voltage of the lamp, the voltage across C_2 when the lamp strikes is V_sC_1/C_2 volts. If V_p is the peak voltage to be measured then, when the lamp just strikes,

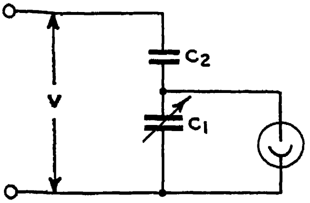


FIG. 16–10

$$V_p = V_s + V_s\frac{C_1}{C_2}$$

$$= V_s\left(1 + \frac{C_1}{C_2}\right)$$

Electronic Ammeter

In general no particular difficulties exist in measuring moderate or large alternating currents at commercial frequencies. For the measurement of small currents at high frequencies, however, facilities are somewhat limited. For such measurements there exist thermal instruments, the oscillograph, metal-rectifier instruments, and the electronic ammeter to be now described. This last instrument depends for its operation on the variation of anode current with cathode temperature. The relationship between these quantities for a state of saturation was derived in Chapter IV, and is expressed by (4–7). Now the cathode temperature is a function of the heater current and resistance, the cathode rate of energy dissipation being I^2R, where I and R are respectively the heater current and resistance. If the heat is dissipated solely by radiation, then the rate of energy loss will be governed by the Stefan-Boltzmann Law and we shall have

$$I^2R = \sigma(T_c^4 - T_0^4) \quad . \quad . \quad . \quad (16\text{–}2)$$

where T_c is the cathode temperature and T_0 the ambient temperature. Hence if the left-hand member of (16–2) is increased by an increase in current so also will be T_c and the saturation current of the anode circuit.

A practical form of ammeter is shown by Fig. 16–11. A diode valve is employed as one arm of a bridge network, the cathode being fed through a choke and a variable resistance R_4. The voltage applied to the heater is lower than its rated figure in order to produce easily a saturation current. The bridge is balanced for d.c. conditions, the galvanometer *G* giving no deflexion in the absence of an alternating current. The latter is applied by the cathode via a condenser, the purpose of the condenser being to isolate the a.c. from the cathode heater supply. To prevent the current under measurement

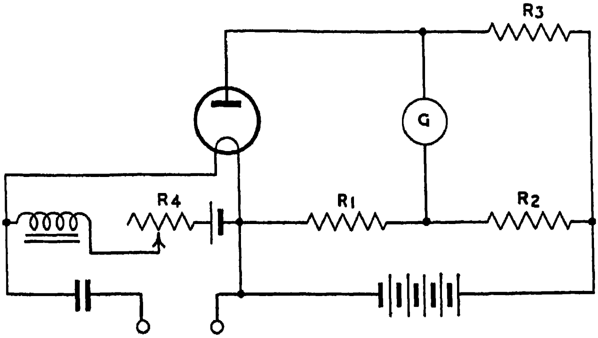


FIG. 16–11

from by-passing the cathode, the iron-cored choke is employed. If $I_{a.c.}$ is the value of the alternating current to be measured and $I_{d.c.}$ the cathode current due to the d.c. supply *E*, then the cathode

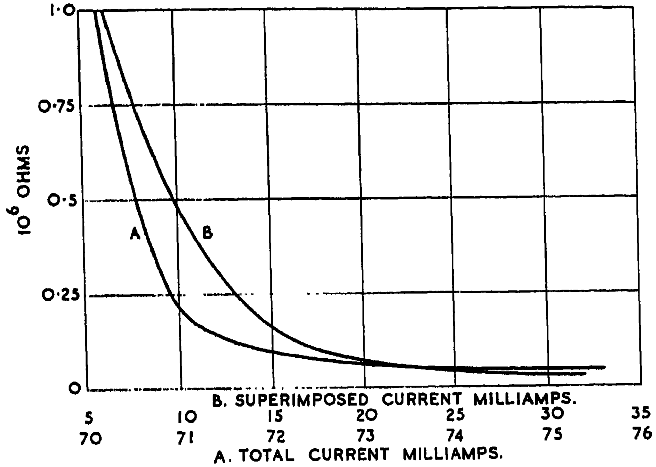


FIG. 16·12

heater current is given by $\sqrt{(I_{d.c.}^2 + I_{a.c.}^2)}$. The presence of $I_{a.c.}$ increases the cathode temperature with a resulting increase in anode saturation current. This unbalances the bridge and the value of the a.c. current may be read on G when the latter is suitably calibrated.

The effect of superimposing an alternating current on the d.c. filament current of a PM6 valve is shown by Fig. 16–12, in which anode resistance is plotted against superimposed a.c. current and total filament current. The normal filament current of this valve is 0·1 amp. but, to produce saturation effects, this was reduced to 0·07 amp. The anode voltage was 116 volts. Fig. 16–13 gives the relation between superimposed a.c. current and deflexion (with a 0–100 microamp. galvanometer) with the same valve employed in the circuit of Fig. 16–11. The values of the various components are as follows—

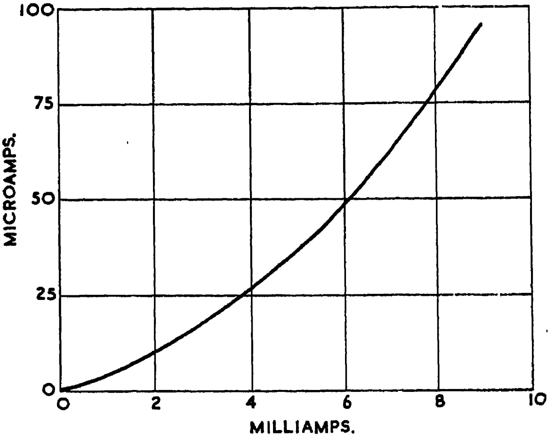


FIG. 16–13

R_1	100,000 ohms	R_3	50,000 ohms
R_2	50,000 ,,	R_4	30 ,,
Choke	50 henrys		

It will be noted that (as with most thermal instruments) the scale tends to be closed at the origin, opening out as the deflexion increases.

ZERO DRIFT

The ammeter of Fig. 16–11 may suffer from the disadvantage of zero drift due to the valve cathode current not remaining strictly constant. This tends to limit the lower range of the instrument because slight variations in the d.c. current may be as large as the a.c. current undergoing measurement. The purpose of the resistance

R_4 is for zero adjustment, but tends to be inadequate where it is desired to measure very small currents.

In order to obviate zero drift the arrangement of Fig. 16–14 may be employed. Here the resistance R_1 is replaced by a diode valve, which must be matched to the operating valve. It will be noted that the filaments of the two valves are in series, so that a change of filament current affects each equally. Assuming the valves to be matched, a change in filament current, common to both, will so affect the internal resistances of the valves that the state of balance of the bridge will tend to remain unaffected. Furthermore, the compensating valve increases the sensitivity of the voltmeter. Thus, when an a.c. current is passed through the cathode of the operating valve, its internal resistance falls and the cathode resistance increases. The latter effect reduces the value of the d.c. current through the cathode of the compensating valve, with the result that its internal resistance rises, further unbalancing the bridge.

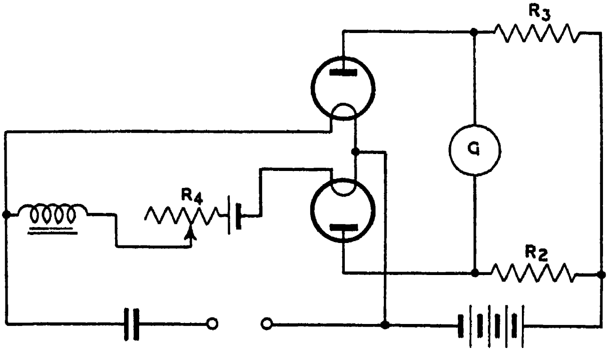


FIG. 16–14

The ammeter outlined by Fig. 16–11 is capable of accurate operation over a wide frequency range, the principal limiting factor being the choke. At low frequencies the reactance of this will fall and a proportion of the current under measurement will tend to pass through the choke rather than through the cathode. Hence low readings may be experienced as the frequency falls. At high frequencies the self-capacity of the choke will tend to lower its reactance, so, again, low readings may result. However, by suitable choke design it appears that satisfactory results may be obtained over a frequency range varying from 25 to 10^6 cycles per second. In addition to the wide frequency range, the instrument is also independent of waveform, as the operation arises from the thermal effect of the current.

Calibration of the ammeter may be carried out by comparison with another instrument at 50 cycles per second. If desired, a number of ranges may be obtained by the use of parallel and series resistances in conjunction with the galvanometer.

The Electronic Wattmeter

The advantages possessed by the electronic wattmeter are similar to those of the voltmeter, i.e. small power consumption and negligible frequency error. In addition it has a relatively high sensitivity compared with the normal type of wattmeter. One type of wattmeter is indicated by Fig. 16–15. Fundamentally it consists of two vacuum valves of similar characteristics and a differential galvanometer arranged as shown, *RR* are identical non-inductive resistances, and R_s a non-inductive potential divider. The galvanometer, *G*, indicates the difference of the anode currents of the two valves and it will be appreciated that a difference is brought about by changes in grid bias due to the load current. The resistance

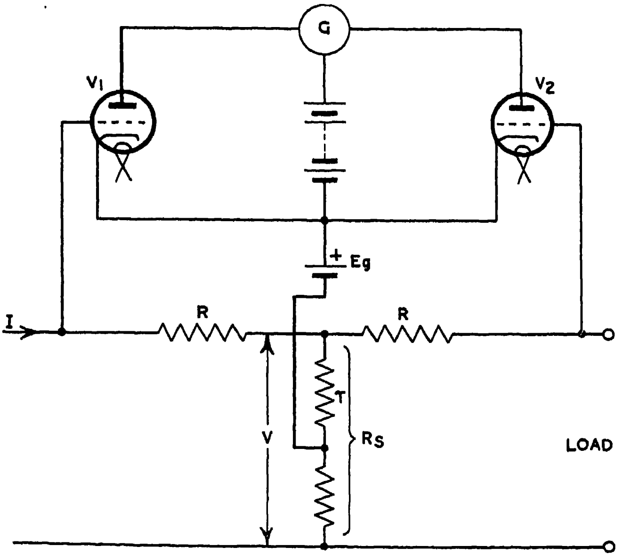


Fig. 16–15

R_s is sufficiently high to make the current flowing in it inappreciable compared with the load current so that when the latter equals I, the voltage across each of RR is IR. The voltage between the junction of RR and the tapping point of R_s is rV/R_s, and hence the grid voltages (apart from the biasing voltage E_g) of the two valves V_1 and V_2 are

$$\left(\frac{rV}{R_s} + RI\right) \text{ and } \left(\frac{rV}{R_s} - RI\right) \text{ respectively}$$

When $I = 0$, these voltages are identical and hence the grids are similarly biased, with the result that the galvanometer reading is zero.

Now assuming the valve static characteristics can be expressed by a square law, the anode current of either valve may be written

$$\begin{aligned} I_a &= a[E_a + \mu(E_g + V_g) + c]^2 \\ &= a\mu^2\left(\frac{E_a + c}{\mu} + E_g + V_g\right)^2 \\ &= a(b + V_g)^2 \end{aligned}$$

where a, μ, b, and c are constants and V_g is a change of grid voltage on either valve. The individual anode currents of either valve are accordingly given by

$$I_1 = a\left(b + \frac{rV}{R_s} + RI\right)^2 \quad . \quad . \quad . \quad (16\text{–}3)$$

$$I_2 = a\left(b + \frac{rV}{R_s} - RI\right)^2 \quad . \quad . \quad . \quad (16\text{–}4)$$

The difference of (16–3) and (16–4) is the galvanometer current I_g and is given by

$$I_g = 4abRI + 4a\frac{rR}{R_s} VI \quad . \quad . \quad (16\text{–}5)$$

If V and I are alternating quantities, then the mean value of (16–5) is

$$I_g = 4a\frac{rR}{R_s}\overline{VI}$$

and the galvanometer reading is proportional to the mean value of the power.

It will be noted that the voltage across the load is less than V

by an amount IR. Hence the power load, W, is the mean value of $(VI - I^2R)$ and we have

$$W = \frac{R_s}{4arR} I_g - I^2R \quad . \quad . \quad . \quad (16\text{–}6)$$

Generally, however, the second term of (16–6) is negligible, with the result that the instrument is direct reading.

EFFECT OF DIFFERENT VALVE CHARACTERISTICS

In the event of valve characteristics differing, the galvanometer no longer gives a reading proportional to the mean power of the load. For example, suppose the two valve currents may be represented by the following equations—

$$I_1 = a_1\left(b_1 + \frac{rV}{R_s} + RI\right)^2$$

$$I_2 = a_2\left(b_2 + \frac{rV}{R_s} - RI\right)^2$$

Then

$$I_g = a_1b_1^2 - a_2b_2^2 + 2(a_1b_1 - a_2b_2)\frac{rV}{R_s} + 2(a_1b_1 + a_2b_2)RI + (a_1 - a_2)\frac{r^2V^2}{R_s^2} + (a_1 - a_2)R^2I^2 + 2(a_1 + a_2)\frac{rR}{R_s}VI$$

and the average value of this is

$$a_1b_1^2 - a_2b_2^2 + (a_1 - a_2)\left(\frac{r^2}{R_s^2}\overline{V^2} + R^2\overline{I^2}\right) + 2(a_1 + a_2)\frac{rR}{R_s}\overline{VI}$$

The first term of this expression may be regarded as a zero error, which can, if necessary, be eliminated by adjustment of the galvanometer. The second term is proportional to the watts lost in r and R, while the third term is proportional to the mean power as before. By making R and r/R small, it is possible, except for small powers, to make the second term negligible compared with the last, when

$$I_g = 2(a_1 + a_2)\frac{rR}{R_s}\overline{VI}$$

Where difference in the characteristics cannot be reduced to a negligible quantity in the foregoing manner, an adjustment for the different slopes may be made in the manner indicated by Fig. 16–16. Here a high resistance is shunted across an ordinary galvanometer and a variable tapping point taken to the battery connexion. In this case the galvanometer current is

$$I_g = \frac{I_1 s_1 - I_2 s_2}{s_1 + s_2 + R_g} \quad . \quad . \quad . \quad (16\text{–}7)$$

of which the numerator is

$$s_1a_1b_1^2 - s_2a_2b_2^2 + 2(s_1a_1b_1 - s_2a_2b_2)\frac{rV}{R_s} + 2(s_1a_1b_1 + s_2a_2b_2)RI$$
$$+ (s_1a_1 - s_2a_2)\frac{r^2V^2}{R_s^2} + (s_1a_1 - s_2a_2)R^2I^2 + 2(s_1a_1 + s_2a_2)\frac{rR}{R_s}VI$$

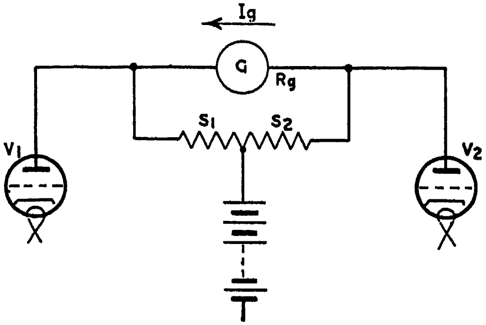


FIG. 16–16

Now if the tapping point is adjusted so that

$$a_1s_1 = a_2s_2 = K, \text{ say},$$

then

$$I_g = \frac{K}{s_1 + s_2 + R_g}\left[b_1^2 - b_2^2 + 2(b_1^2 - b_2^2)\frac{rV}{R_s} + 2(b_1 + b_2)RI + 4\frac{rR}{R_s}VI\right]$$

the average value of which is

$$\bar{I}_g = \frac{K(b_1^2 - b_2^2)}{s_1 + s_2 + R_g} + \frac{4K}{s_1 + s_2 + R_g}\cdot\frac{rR}{R_s}\overline{VI}$$

The first term of this may be eliminated in the manner previously described when the galvanometer will indicate the mean power as with valves with identical characteristics.

In order to vary the range of the wattmeter, it is only necessary to change RR and the tapping point on R_s. However, when it is desired to measure low powers with small currents in the series resistances RR, the latter may have to be of such value as to cause the I^2R loss to be of the same order of magnitude as the power undergoing measurement. Correction may, of course, be made for this as indicated by (16–6), but may be avoided by altering the equality of the two resistances R. Calling these R_1 and R_2, we have

$$I_1 = a_1\left(b_1 + \frac{rV}{R_s} + R_1I\right)^2$$

$$I_2 = a_2\left(b_2 + \frac{rV}{R_s} - R_2I\right)^2$$

and the numerator of (16–7) now becomes

$$s_1a_1b_1^2 - s_2a_2b_2^2 + 2(s_1a_1b_1 - s_2a_2b_2)\frac{rV}{R_s} + 2(s_1a_1b_1R_1 + s_2a_2b_2R_2)I$$
$$+ (s_1a_1 - s_2a_2)\frac{r^2V^2}{R_s^2} + (s_1a_1R_1^2 - s_2a_2R_2^2)I^2 + 2(s_1a_1R_1 + s_2a_2R_2)\frac{r}{R_s}VI$$

Again, making $a_1s_1 = a_2s_2 = K$

we have

$$I_g = \frac{K}{s_1 + s_2 + R_g}\Big[(b_1^2 - b_2^2) + 2(b_1 - b_2)\frac{rV}{R_s} + 2(b_1R_1 + b_2R_2)I$$
$$+ (R_1^2 - R_2^2)I^2 + 2(R_1 + R_2)\frac{r}{R_s}VI\Big]$$

The mean value of the second and third terms is zero, and eliminating the first term by zero adjustment of the galvanometer

$$I_g = \frac{K}{s_1 + s_2 + R_g}\left[(R_1^2 - R_2^2)\overline{I}^2 + 2(R_1 + R_2)\frac{r}{R_s}\overline{VI}\right]$$
$$= \frac{K}{s_1 + s_2 + R_g}\cdot\frac{2(R_1 + R_2)r}{R_s}I\left[V - \frac{R_2^2 - R_1^2}{R_2 + R_1}\cdot\frac{R_s}{2r}I\right]$$

Now the load voltage is $V - R_2I$ and, hence, if

$$R_2 = \frac{R_2^2 - R_1^2}{R_2 + R_1}\cdot\frac{R_s}{2r} \quad . \quad . \quad . \quad (16\text{–}8)$$

the galvanometer will read the load power direct without correction. The relation of R_1 to R_2 is found from (16–8), viz.

$$R_1 = \left(1 - \frac{2r}{R_s}\right)R_2$$

The relatively high sensitivity of the foregoing form of wattmeter has already been mentioned, and it may be added that an instrument has been constructed for a power as low as one milliwatt full-scale deflexion.

Electronic Ohmmeter

An electronic ohmmeter which is particularly suitable for the measurement of high resistances, such as insulation resistances, is shown by Fig. 16–17. It will be noted that the indicating device consists of a cathode-ray tuning indicator.* The resistance undergoing measurement is connected between the terminals AB, and the resistance current then flows through R_1 thus altering the grid potential and shadow angle. If the value of the resistance undergoing measurement is denoted by X, and E is the voltage between AB, the current through X is $E/(X + R_1)$. Thus, the effect of this current through R_1 is to decrease the potential difference between grid and cathode of the tuning indicator, consequently increasing the shadow angle. Denoting the change in grid potential by V_{gc}, we have

$$V_{gc} = \frac{ER_1}{X + R_1}$$

and

$$X = \frac{R_1(E - V_{gc})}{V_{gc}}$$

Thus, for a given change in grid potential and shadow angle, X is proportional to R_1 and the value of X may be determined from that of R_1.

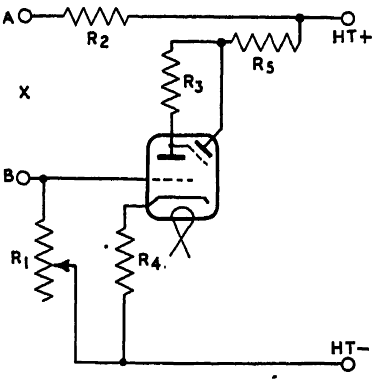


Fig. 16–17

Normally the ohmmeter shown is only suitable for the measurement of large resistances because the voltage between AB cannot be less than the target voltage, i.e. about 200 volts. Where the instrument is employed for the determination of insulation resistance the working voltage will generally be at least 500 volts. In this case the resistance R_5 is employed to obtain the correct target voltage. To avoid short-circuiting the h.t. supply, should X be abnormally low, R_2 may be incorporated.

In order to operate the ohmmeter, R_1 is initially set at zero and, with AB on open circuit, R_4 should be such that the shadow angle is, say, 15° corresponding to a grid potential V_{g1}. X is now connected and R_1 adjusted until a shadow angle of, say, 45° is obtained, corresponding to a grid potential of V_{g2}. Then $V_{gc} = V_{g1} - V_{g2}$ and

$$X = \frac{R_1[E - (V_{g1} - V_{g2})]}{V_{g1} - V_{g2}}$$

---

* See p. 682.

If the same shadow angle (45°) is always employed, then $(V_{g1} - V_{g2})$ may be determined for all time and, knowing E, X may be found from the value of R_1. A better plan than this, however, is to connect known resistances across AB, adjust R_1 for the same shadow angle in every case, and calibrate the dial of R_1 in terms of X. Alternatively, the dial of R_1 may be divided into a number of equally spaced divisions (say, 100) and a calibration curve drawn. The latter plan is better as the calculation of X from $(V_{g1} - V_{g2})$, E, and R_1 is complicated because, due to the regulation of the h.t. supply, E will be different for every value of X, especially if R_2 is incorporated.

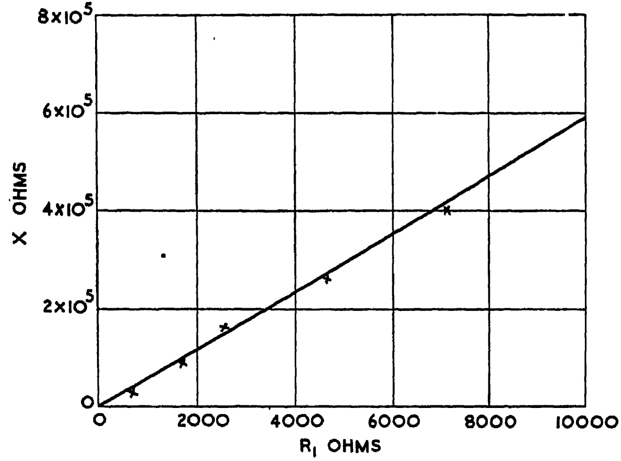


FIG. 16–18

The range of the instrument may, of course, be extended by employing a number of resistances for R_1, selecting the appropriate one by means of a switch.

An example of a typical ohmmeter for the measurement of resistances up to 600,000 ohms may be given by the following details. The tuning indicator employed is a Tungsram 6U5 with a target voltage of 250 volts. R_4 is 5400 ohms and, with a supply current of 2mA, this gives a bias of 10·8 volts and a shadow angle of about 20°. R_1 is adjusted for a 45° shadow angle and in these circumstances the grid bias is 6 volts. Thus the change in grid potential is 4·8 volts and

$$X = 52R_1 \text{ approx.}$$

The values of R_3 and R_2 are respectively 1 megohm and 2500 ohms, while the maximum value of R_1 is 10,000 ohms. Fig. 16–18 shows the experimentally determined relation between X and R_1 which, it will be noted, is linear.

Electronic Harmonic Analysis

A form of harmonic analyser, electronic in character, is shown by Fig. 16–19. An instrument similar to a dynamometer wattmeter is employed as shown at *E*. The moving coil of this is connected to the voltage to be analysed and thus carries a current proportional to this voltage. The fixed coil of the instrument carries a current, called the analysing current, the frequency of which can be made equal to that of the fundamental or any harmonic in the wave undergoing analysis. The frequency of the analysing current is produced by means of a synchronously driven contact disc, the latter possessing concentric rings of contacts, each contact alternating with an insulating segment. From the arrangement of the grid bias it will be appreciated that, as the disc revolves, alternate positive and negative bias voltages are impressed on the triode grid for equal periods. Thus, if there is one contact on the disc occupying one-half of its periphery and the synchronous motor runs at 3000 r.p.m., a 50-cycle *rectangular* voltage wave will be impressed between grid and cathode. By increasing the number of contacts evidently any desired frequency can be produced.

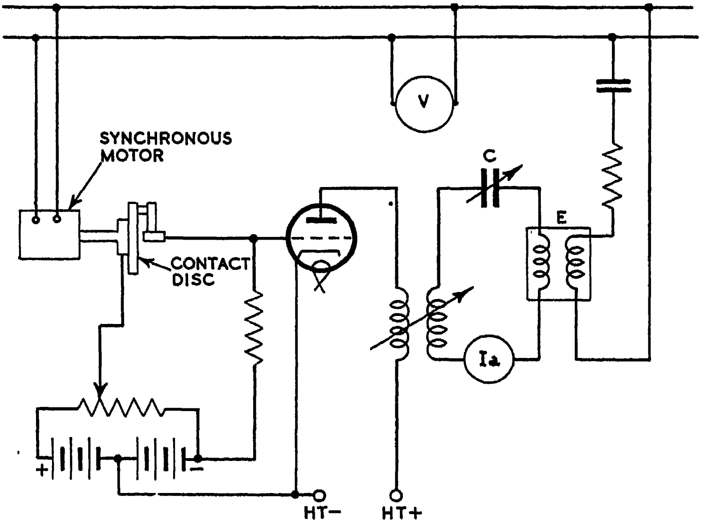


FIG. 16–19

The rectangular voltage wave on the grid will, of course, produce an alternating component in the anode current of the same frequency, but, due to the inductance of the variable-coupling transformer, this component will have a closer resemblance to a sine wave. The waveform in the transformer secondary winding is further improved by tuning the secondary circuit to the harmonic frequency by means of the variable condenser C.

Now it is well known that a steady deflexion of a dynamometer instrument results only when the currents in the fixed and moving coils are of the same frequency. Thus, if the analysing current is

$$i_a = I_a \sin (n\theta + \phi_a)$$

and the current in the moving coil due to the nth harmonic is

$$i = I_m \sin (n\theta + \phi_m),$$

the instantaneous torque is given by

$$T \propto i i_a$$

The mean torque is

$$T_m = \frac{K}{2\pi}\int_0^{2\pi} i i_a d\theta$$

$$= \frac{K}{2\pi}\int_0^{2\pi} I_a \sin (n\theta + \phi_a) I_m \sin (n\theta + \phi_m) d\theta$$

$$= K I_a I_m \cos (\phi_m - \phi_a)$$

As the deflexion is proportional to the torque, we may write

$$D = K_1 I_a I_m \cos (\phi_m - \phi_a)$$

where D is the deflexion. The phase of the analysing current is adjusted by the contact arm of the rotating disc until $(\phi_m - \phi_a) = 0$. At this point D is a maximum and

$$I_m = D_m / K_1 I_a$$

where D_m is the maximum deflexion obtained. To obtain the value of the harmonic voltage, V_n, we have

$$V_n = \frac{I_m}{\sqrt{2}}\sqrt{R^2 + \left(npL - \frac{1}{npC_1}\right)^2}$$

where $p = 2\pi f$, f being the frequency of the fundamental, and L, C_1, and R the inductance, capacitance, and resistance of the moving-coil circuit.

Where the phase relationship of the harmonics is not needed, harmonic analysis may be effected by means of an electronic voltmeter. The voltage to be analysed is applied in series with a local oscillator to the grid and cathode of the valve, in the anode circuit of which is the usual milli- or micro-ammeter. The frequency of the oscillator is adjusted to the frequency of the harmonic to be measured and, as this adjustment is made, the milliammeter indicator will oscillate slowly in response to the heterodyne frequency difference. The oscillation frequency is, of course, the difference between the harmonic and local oscillator frequencies, and when, by adjustment, this difference is made sufficiently small, the amplitude of the anode current variation may be read directly. From this amplitude and that of the local oscillation (as found from the voltmeter when the input is zero) the amplitude of the harmonic may be calculated.

Assuming that the grid-voltage/anode-current characteristic follows a quadratic law we have

$$i = a_1 v_g + a_2 v_g^2 \quad . \quad . \quad . \quad (16\text{–}9)$$

provided that the quiescent current has been balanced out. If v_0 is the voltage of the local oscillation and v_1 that to be analysed, then (16–9) becomes

$$i = a_1 v_0 + a_1 v_1 + a_2 v_0^2 + a_2 v_1^2 + 2a_2 v_0 v_1 \quad . \quad (16\text{–}10)$$

Now let $v_0 = V_0 \cos \phi_0$

and $\quad v_1 = V_1 \cos \phi_1 + V_2 \cos 2\phi_1 + V_3 \cos 3\phi_1 + \ldots . \quad (16\text{–}11)$

This, of course, presumes that the local oscillator waveform is sinusoidal and that the phase angles of the harmonics are zero. The modification necessary when the latter is not true will be shown later. Substituting (16–11) in (16–10)—

$$\begin{aligned} i = {} & a_1 V_0 \cos \phi_0 + a_1 V_1 \cos \phi_1 + a_1 V_2 \cos 2\phi_1 + \ldots \\ & a_2 V_0^2 \cos^2 \phi_0 + a_2 V_1^2 \cos^2 \phi_1 + a_2 V_2^2 \cos^2 \phi_1 + \ldots \\ & 2a_2 V_0 V_1 \cos \phi_0 \cos \phi_1 + 2a_2 V_0 V_2 \cos \phi_0 \cos 2\phi_1 \ldots \\ & 2a_2 V_1 V_2 \cos \phi_1 \cos 2\phi_1 + 2a_2 V_1 V_3 \cos \phi_1 \cos 3\phi_1 + \ldots \\ & \ldots + \ldots + \ldots + \ldots \end{aligned}$$

Grouping terms—

$$\begin{aligned} i = {} & a_1 \sum_0^n V_n \cos n\phi_1 + a_2 \sum_0^n V_n^2 \cos^2 n\phi_1 \\ & + 2a_2 \sum_1^n V_0 V_n \cos \phi_0 \cos n\phi_1 \quad . \quad . \quad . \quad (16\text{–}12) \\ & + a_2 \sum_1^n \sum_1^m V_n V_m \cos n\phi_1 \cos m\phi_1 \\ & \qquad n \neq m \end{aligned}$$

it being understood that when $n = 0$ in the first two terms, $\cos n\phi_1$ is replaced by $\cos \phi_0$.

Changing the trigonometrical terms we may write

$$i = a_1\sum_0^n V_n \cos n\phi_1 + \frac{a_2}{2}\sum_0^n V_n^2 - \frac{a_2}{2}\sum_0^n V_n^2 \cos 2n\phi_1$$

$$+ a_2\sum_1^n V_0 V_n \cos(\phi_0 + n\phi_1) + a_2\sum_1^n V_0 V_n \cos(\phi_0 - n\phi_1)$$

$$+ a_2\sum_1^n\sum_1^m V_n V_m \cos(n\phi_1 + m\phi_1) + a_2\sum_1^n\sum_1^m V_n V_m \cos(n\phi_1 - m\phi_1) \quad (16\text{–}13)$$

$$n \neq m$$

Now the milliammeter will indicate the steady components of (16–13), plus those of such frequency that their effect is easily readable, i.e.

$$\frac{a_2}{2}\sum_0^n V_n^2 + a_2\sum_1^n V_0 V_n \cos(\phi_0 - n\phi_1) \quad . \quad . \quad (16\text{–}14)$$

Now by adjusting ϕ_0 to be approximately equal to $n\phi_1$, the beat frequency may be made so small that its effect on the meter is easily discernible. Other beat frequencies, of course, coexist, but, with the adjustment effected as described, their effect cannot be detected. Referring to (16–14), the swing of the milliammeter indicator is

$$I_b = 2a_2 V_0 V_n$$

and

$$V_n = \frac{I_b}{2a_2 V_0}$$

which thus determines the amplitude of the nth harmonic. I_b is read directly, while V_0 is found by employing the analyser as a voltmeter to read the oscillator voltage in the absence of that undergoing analysis. a_2 is found by employing the instrument as a voltmeter and applying a known voltage. If V is the peak value of this voltage

$$i = \frac{a_2}{2} V^2 \quad . \quad . \quad . \quad . \quad (16\text{–}15)$$

where i is the anode current resulting from V. Thus (16–15) determines a_2.

Where the phase angles are different from zero we must write

$$v_1 = V_1 \cos(\phi_1 + \psi_1) + V_2 \cos(2\phi_1 + \psi_2) + \ldots$$

in which case the reading of the milliammeter will be

$$\frac{a_2}{2}\sum_{1}^{n} V_n^{\,2} + a_2\sum_{1}^{n} V_0 V_n \cos\,(\phi_0 - n\phi_1 - \psi_n) \qquad . \quad (16\text{–}16)$$

i.e. it will not differ from that indicated by (16–14).

BIBLIOGRAPHY

"Electrical Non-Destructive Testing of Materials," G. R. Polgreen and G. M. Tomlin, *Electronic Engineering*, April, 1946, p. 100.

"A New Fuel Gauge for Aircraft," *The Engineer*, 1st Mar., 1946, p. 204.

"An Electrical Moisture Meter," C. F. Brockelsby, *J. Sci. Inst.*, Dec., 1945, p. 243.

"An Electrical Moisture Meter," L. Hartshorn and W. Wilson, *J.I.E.E.*, Oct., 1945, Pt. II, p. 403.

"A Valve Ammeter," H. E. M. Barlow, *J.I.E.E* Vol. LXXVII, p. 612.

"A Very High Impedance R.M.S. Voltmeter for Iron Testing," D. C. Gall and F. C. Widdis, *J. Sci. Inst.*, Dec., 1946, p. 287.

CHAPTER XVII

CATHODE-RAY TUBES AND ASSOCIATED CIRCUITS

THE cathode-ray tube may be justly regarded as one of the most important applications of electron optics and dynamics that exists. In what is, perhaps, its best-known form it consists, fundamentally, of a glass tube in which are housed a cathode, an anode, or gun, an electron optical system, and two or more deflecting plates. This arrangement is shown by Fig. 17–1, and in this example the domed end of the tube is coated with a fluorescent powder in order that a trace may be formed thereon by an electron beam. The electrons are produced thermionically, as described in Chapter IV, and, under the influence of the positive potential of the anode, pass up the tube towards the fluorescent screen. In so doing they pass through a small hole in the anode, which tends to restrict the electrons to a narrow beam, in order that a small patch or spot of light may result from the impinging of the electrons on the screen. Following the anode is some form of focusing system by means of which the beam is brought to a fine focus at the recording surface. Between the latter and the lens are deflecting plates or coils in order that the phenomena under consideration may deflect the beam and so record its characteristics in a manner which may be readily interpreted.

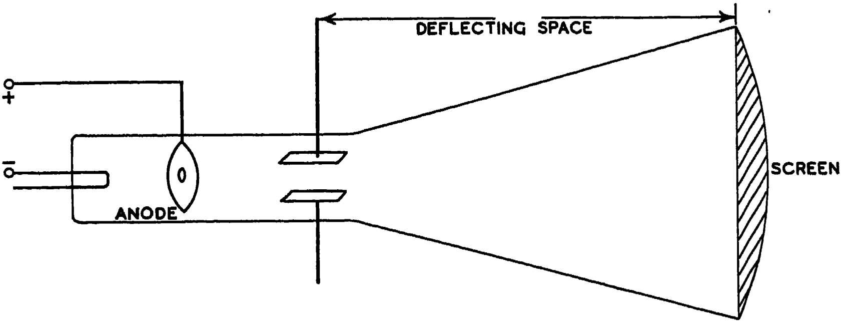


FIG. 17–1

Cathode tubes may be classified under three headings:—

1. Low-voltage tubes, which are of the gas-focused type. With these the accelerating voltage seldom exceeds 1000 volts.

2. Medium-voltage tubes employing accelerating voltages

between 1000 and 10,000 volts. These are of the high-vacuum type and are focused by an electrostatic lens as described in Chapter VI.

3. High-voltage tubes in which the accelerating potential is from 10 kV to 80 kV. Such tubes usually employ a cold cathode and are focused by a magnetic lens.

Gas-focused Tubes

Considering first the gas-focused tube, as stated above, the beam is restricted in cross-section by a small aperture in the anode before focusing occurs. This arrangement, however, is somewhat inefficient, in that the majority of electrons are collected by the anode, only a small proportion passing up the tube to the screen. This state of affairs is remedied by surrounding the cathode with a small metal tube known as a Wehnelt cylinder, or, more commonly, as a shield or grid. This is negatively biased to about 50 volts and the resulting electrostatic field has a converging effect on the initially diverging electrons as they leave the cathode. The result is that a much larger proportion of the electrons pass into the deflecting space which, of course, leads to a greater concentration of electrons in the beam. Optically, the shield is equivalent to a convex lens of short focal length, which produces an image of the cathode just in front of the anode aperture.

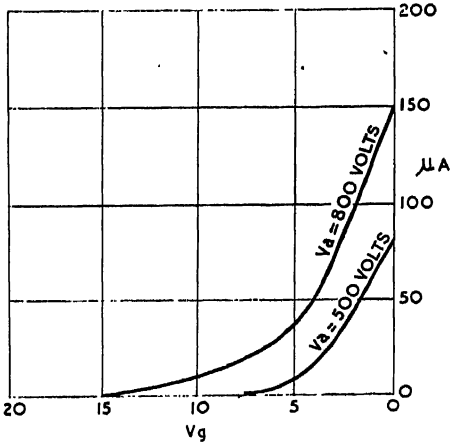


FIG. 17-2

In many respects the action of the shield is similar to that of the grid in a triode, for variation of the value of the biasing voltage varies the anode current. If the bias is continually increased, a value is reached at which the anode current decreases to the extent that the spot vanishes from the fluorescent screen. Evidently the brilliance is also controllable by the shield bias and, in practice, cathode-ray tubes are provided with continuously variable biasing arrangements for controlling the brilliance and sharpness of the spot and for putting it out of existence when desired.

The relationship between anode current and shield voltage for a typical case is shown by Fig. 17–2, from which the similarity to the static characteristic of an ordinary triode valve may be noted.

EFFECT OF GAS FILLING

From the nature of the deflector plate circuit it might appear that the cathode-ray tube imposes only a small capacitive load on the circuit producing the phenomenon undergoing examination. In gas-filled tubes, however, a resistive loading is also imposed, due to the flow of electrons and ions to the deflector plates when a potential difference is applied to the latter. Where the impedance of the source of supply is low, this is unimportant, but should it be high (say, of the order of 10^7 ohms), distortion of the applied waveform is likely to occur. In general, it may be said that whenever the supply impedance is of the same order of magnitude as that of the plate circuit, some distortion will result. Due allowance could be readily made for this if the plate circuit impedance were constant. Actually this impedance is far from constant, as will be apparent from Fig. 17–3, which shows plate resistance plotted as a function of plate potential in a typical case. It will be noted that the resistance is a minimum at zero voltage, i.e. when the beam is undeflected. It follows that with a high-impedance supply the tube is less sensitive near the origin than with more remote positions of the beam.

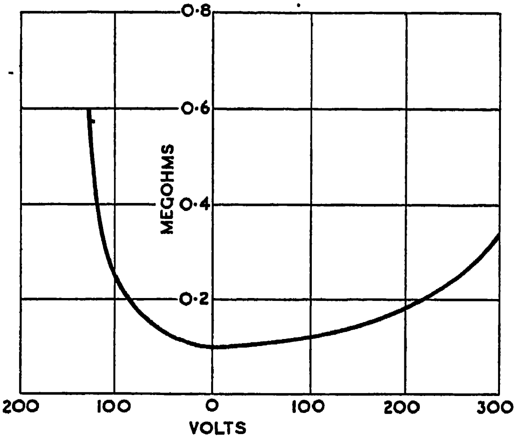


FIG. 17–3

In order to minimize the variation of plate impedance it is customary to connect fixed resistances across the deflector plates in the manner indicated by Fig. 17–33. This also ensures that the tube is not operated on an open circuit, for in such circumstances charges would be acquired by the plates leading to spurious results.

Although it has been stated above that the finite impedance of the plate circuit is due to the presence of electrons and ions, it appears that it is principally electrons which are responsible for this condition. The electrons are secondary in character and are dislodged from the screen by the impacting beam electrons. The secondaries return to the anode, but to reach this must pass the deflector plates, where they will be either attracted or repelled by the charges thereon. This results in a relatively high current for positive plate potentials as shown by Fig. 17–4, Curve 1. The current may be reduced to a low figure by extending the anode, in the form of a tube, beyond the deflector plates. This intercepts and collects the secondary electrons, resulting in the low plate current shown by Curve 2, Fig. 17–4. With this structure the plate impedance may be as high as 2×10^7 ohms.

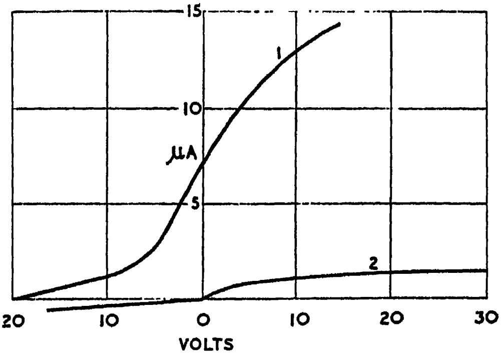


FIG. 17–4

ORIGIN DISTORTION

A further form of distortion in gas-filled tubes is known as *origin distortion*. When a deflecting potential is applied to the plates, electrons and ions tend to collect in their neighbourhood, thus forming space-charges. Because of their lower mobility, the interplate space is chiefly occupied by positive ions. The resulting positive space-charge distorts the electrostatic field, causing a cathode fall of potential at whichever plate happens to be negative. An effect of this is to absorb part of the potential applied to the plates to balance the space-charge drop, with the result that the tube is less sensitive to low plate potentials than high ones.

A deflecting system overcoming origin distortion is indicated by Fig. 17–5, and is due to von Ardenne. It will be noted that one of the deflecting plates is equally divided and equal but opposite biasing potentials applied to each half. The plate division is preferably and generally applied to that plate which is at earth potential. If the value of the bias potential is at all times higher than that of the deflecting potential, then the resultant potential of the system is always different from zero and origin distortion will not occur. In practice the bias voltage applied to each half-plate is rather greater than that necessary to deflect the spot to the edge of the fluorescent screen.

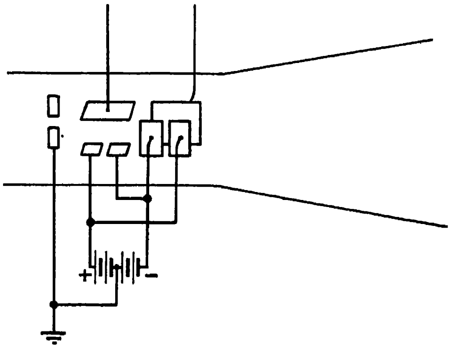


Fig. 17–5

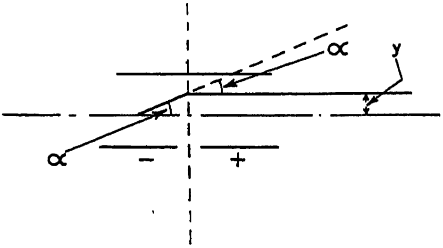


Fig. 17–6

The action of the biasing potentials on the beam in the absence of a deflecting potential may be appreciated with the assistance of Fig. 17–6. On entering the plate space the beam is deflected through the angle α and, but for the influence of the positive bias, would continue along the broken line. However, due to the positively biased half-plate, the beam is subsequently deflected through an angle equal to α, but in the opposite direction. Thus the beam merely undergoes a small lateral displacement y, which is easily corrected if necessary.

HIGH-FREQUENCY DEFOCUSING

An effect experienced with gas-filled tubes is defocusing at high frequencies. This is because high-frequency deflexion of the beam may not give sufficient time for an adequate supply of positive ions to be formed to constitute the core of the beam. The effect may be countered to a limited extent by certain measures, such as increasing the beam current at high frequencies and employing a gas-filling of low molecular weight. However, the frequency limit of the gas-filled tube is about 10^5 cycles per second, and for frequencies beyond this a high-vacuum tube must be employed.

Medium-voltage High-vacuum Tubes

Because of the various limitations of the gas-filled tube, detailed above, high-vacuum tubes are largely employed and are actually a necessity for television purposes. These tubes are evacuated to about 10^{-6} mm. Hg, and focusing is effected by an electrostatic lens of the form described in Chapter VI. Magnetic lens are seldom used because, the tube being short, the stray field would extend over a considerable length of the beam, and because of the d.c. power needed for the coil excitation.

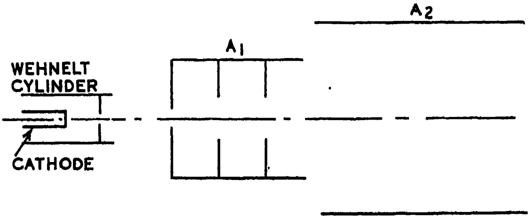


FIG. 17–7

A result of the relatively high voltage employed with vacuum tubes is a reduced sensitivity as compared with the gas-focused tube. Thus, it is desirable from the sensitivity viewpoint that the deflecting system should be near the cathode while for a sharply focused beam the lens should be near the recording surface. However, the lens must precede the deflector plates, and hence a compromise must be effected between high sensitivity and sharp focusing.

The types of lenses usually employed are the two and three-diaphragm and coaxial cylinders. With low voltages the two-diaphragm type is largely employed because of the smaller image it produces due to the axial acceleration of the beam. A typical cylinder lens is shown by Fig. 17–7, the two cylinders A_1 and A_2 being, respectively, referred to as the first and second anodes. It will be noted that the first anode is fitted with a number of baffles, the purpose of these being to collimate the beam. A Wehnelt cylinder is invariably fitted in high-vacuum tubes, this producing an image of the cathode between itself and the lens. The latter then produces a further image of the first one at the screen or recording surface.

VACUUM TUBE DISTORTION

Although the vacuum tube is free from origin distortion and defocusing at high frequencies, it does, however, possess certain shortcomings peculiar to itself. Due to the relatively low sensitivity of such tubes, comparatively high potentials are necessary on the deflector plates. One of each pair of plates is normally connected to the anode (in both gas-filled and vacuum tubes), while the others periodically fluctuate above and below anode potential. Thus, as the beam is deflected, it will undergo an additional acceleration or retardation, this effect being greater the greater the amplitude of the deflexion. The change produced in the axial velocity naturally tends to defocus the beam, the defocusing increasing as the beam approaches the edges of the screen. A remedy is to arrange matters so that a symmetrical deflexion is obtained in which one plate of each pair rises above the anode potential, while the other is falling below it by the same amount. In this manner the mean potential of the plates relative to the anode remains unchanged. Methods of effecting this will be discussed after describing a further form of distortion.

TRAPEZIUM DISTORTION

As previously shown, the sensitivity of a cathode-ray tube depends on the axial accelerating voltage. However, in a high-vacuum tube this voltage is either sensibly increased or decreased by the deflecting voltage and thus the sensitivity is affected according to the magnitude of the voltage on the deflector plates. Thus, with a positive deflecting voltage, the accelerating potential increases and the sensitivity and resulting deflexion are smaller than when the deflecting potential is negative. The term "trapezium distortion" results from the shape of the television raster produced when

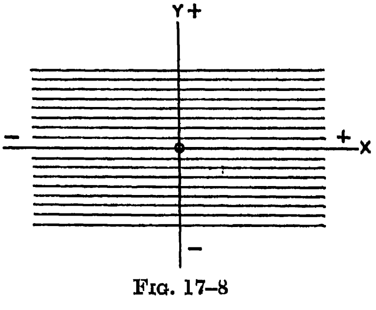


FIG. 17–8

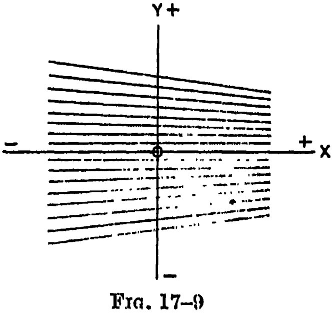


FIG. 17–9

this form of distortion is present. The raster is produced by two simultaneous deflexions, one vertical and the other horizontal, the frequency of the latter being much higher than that of the former. The deflecting voltages are produced by saw-tooth wave generators (see page 560), with the result that, providing distortion is absent, a series of horizontal lines appear on the screen in the manner indicated by Fig. 17–8. However, with trapezium distortion present, if we assume that $+y$ and $+x$ deflexions are produced by positive potentials on the Y- and X-plates, the length of the vertical deflexion will be less to the right of the centre line than to the left of it. The result of this is to produce a raster similar in shape to that shown by Fig. 17–9. Hence the term "trapezium distortion." The picture is symmetrical about $+yO-y$ because of the self-centring property of the tube (see page 562).

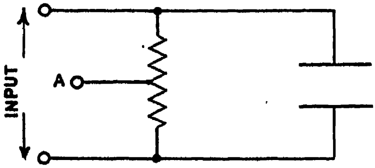


FIG. 17–10

SYMMETRICAL DEFLEXION

Trapezium distortion may be avoided in the same manner as defocusing, i.e. by the employment of symmetrical deflexion. This means that the electrical centre of each plate pair must be connected to the final anode so that the

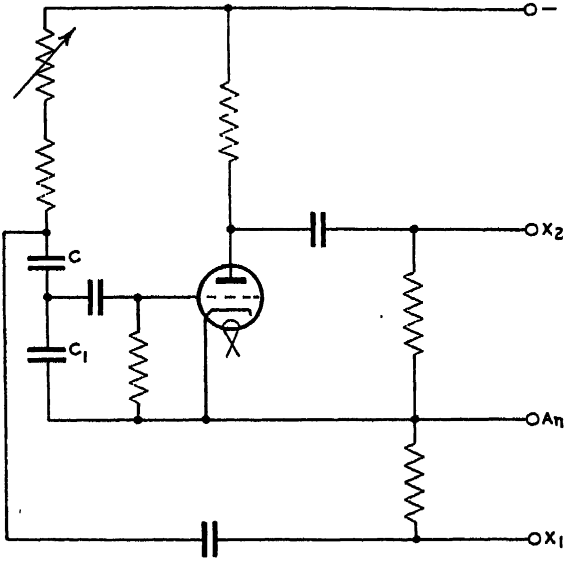


FIG. 17–11

mean potential between this and the plates is zero. One method of effecting this is shown by Fig. 17–10, where the plates are connected by a high resistance the centre of which is connected to the final anode. It is found in practice that satisfactory results are usually obtained if symmetrical deflexion is only applied to the plates nearest the final anode, i.e. the X-plates. A method of application is shown by Fig. 17–11. The time-base condenser is divided into two parts, as shown, and the fluctuating voltage on C_1 applied to the grid of an amplifying valve whose stage gain is M. If C_1 is equal to $(M - 1)C$, then the voltage applied to the grid is $1/M$th of the voltage on $(C + C_1)$. The amplifier output voltage is equal in value to that on the time-base condenser, but, due to the phase-reversing action of the valve, it is opposite in phase. Thus, if the deflector plates are connected as shown by Fig. 17–11, their potentials will at every instant be at equal amounts above and below the anode potential. Thus the mean potential difference of the plates and the anode will be zero.

The High-voltage Tube

For the recording of very high-frequency phenomena, such as transients lasting only a few microseconds, the two foregoing types of tubes cannot be employed. This is because the intensity of the trace is too low to register on a photographic plate. The penetrating power of the beam varies as the square of the accelerating voltage and this, of course, is the reason for the employment of anode voltages between 10 kV and 80 kV for high-speed work. A schematic diagram of a high-speed tube is shown by Fig. 17–12, from which it may be noted that a magnetic lens is employed and also that the photographic plate is included within the tube. The latter feature necessitates that the tube is demountable and continuously connected to vacuum pumps. Because of the difficulties of employing a hot cathode at high voltages (such as positive-ion bombardment), high-voltage tubes invariably possess a cold cathode, the electrons being produced from this by secondary emission caused by positive-ion bombardment. Thus the discharge chamber contains air at a pressure of about 10^{-2} mm. of Hg, in order to permit the production of the necessary ions.

For making visual observations prior to photographing phenomena under examination, it is usual to employ a fluorescent screen. This can be raised from outside when desired, thus exposing the photographic plate to the incident electrons. For preventing the electrons from reaching the recording surface, the beam trap and diaphragm are employed as shown in Fig. 17–12. In many instances high-voltage tubes are constructed of metal, this naturally giving a much more robust structure than when glass is employed. A typical tube is shown by Fig. 17–13, a few dimensions being included to give some idea of the large size of high-voltage tubes.

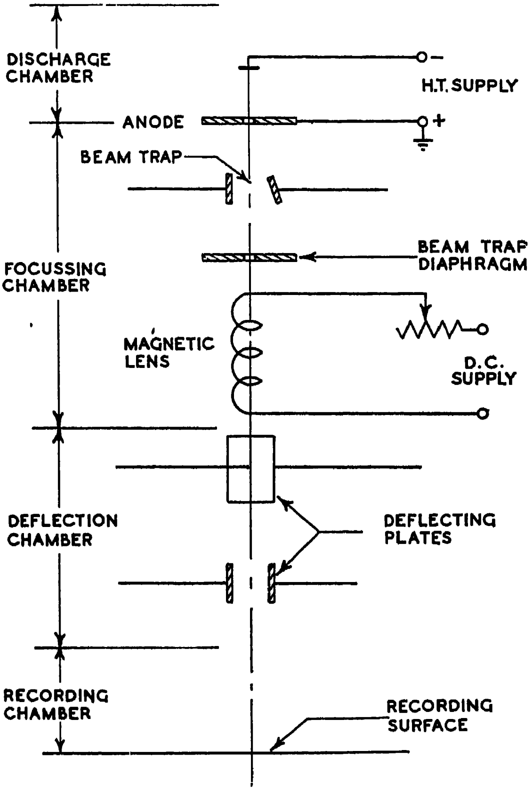


FIG. 17–12

BEAM DEFLEXION

Having formed a satisfactory electron beam by means of a suitable electron-optical system, it is necessary to deflect the beam by the phenomenon under examination in order that a trace may be formed on the recording system at the far end of the cathode-ray tube. This is effected by either an electrostatic or magnetic field, or sometimes by a combination of both. In order to form an electrostatic deflecting field, a pair of plates is fitted inside the tube, as shown by Fig. 17–1, and a deflecting potential V_p applied thereto (Fig. 17–14). The electrostatic field strength between the plates is V_p/d and an electron entering the field experiences a vertical accelerating force (say, downwards)

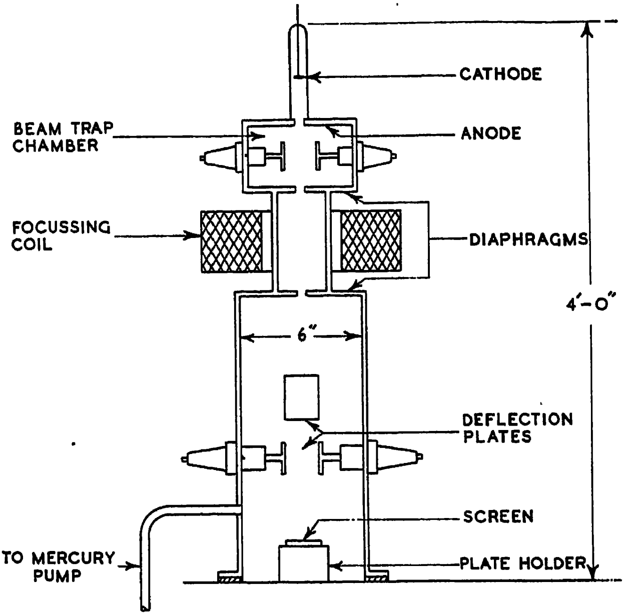


FIG. 17–13

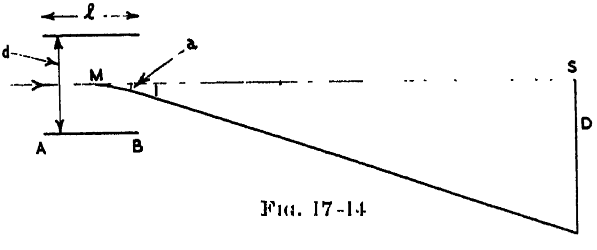


FIG. 17–14

$$eV_p/d$$

The acceleration of the electron is consequently

$$\alpha = eV_p/dm$$

At any time t after entering the plate space, the axial distance travelled by the electron is

$$x = vt \quad . \quad . \quad . \quad . \quad (17\text{–}1)$$

where v is the electron velocity at the moment of entering. In the same time the electron acquires a downward velocity component,

$$v_1 = \alpha t = xeV_p/mvd \quad . \quad . \quad (17\text{–}2)$$

and travels a vertical distance which, from (17–1) and (17–2) is equal to

$$\tfrac{1}{2}eV_p/md \, . \, x^2/v^2 \quad . \quad . \quad . \quad (17\text{–}3)$$

Thus the trajectory of the electron between the plates is a parabola.

From (17–1), the time of flight of the electron between the plates is

$$t = l/v$$

and the downward velocity component at this time is

$$v_1 = leV_p/mvd \quad . \quad . \quad . \quad (17\text{–}4)$$

Referring to Fig. 17–14, the exit velocity of the electron is given by $\tan a = v_1/v$ which, from (17–4), is equal to

$$leV_p/mdv^2$$

But $v^2 = 2eV/m$ and, therefore,

$$\tan a = lV_p/2dV$$

From (17–3), the vertical displacement is

$$\frac{l^2V_p}{4dV} = \frac{l}{2}\cdot\frac{lV_p}{2dV} = \frac{l}{2}\tan a$$

Thus the base of the triangle of Fig. 17–14 is equal to $l/2$, i.e. half the length of the plates. Hence the electron path and beam may be regarded as linear and originating at a point M midway along the plate axis. This means that M is, in effect, a virtual cathode and that electrons appear to emerge from this "cathode" independent of the values of V_p and V. If L is the distance between M and the recording surface, the deflexion of the beam at the latter is

$$D = L \tan a$$

$$= LlV_p/2dV \text{ cm./volt} \quad . \quad . \quad (17\text{–}5)$$

The deflexion on the recording surface is therefore proportional to the plate voltage V_p, and is free from amplitude distortion. The deflexion per volt is defined as the sensitivity of the tube and, from (17–5), is

$$Ll/2dV \quad . \quad . \quad . \quad . \quad (17\text{–}6)$$

The deflexion factor of the tube is the reciprocal of the sensitivity, i.e.

$$\frac{2dV}{Ll}$$

and gives the potential necessary to cause unit deflexion at the screen surface.

It will be noted that the sensitivity of a cathode-ray tube is inversely proportional to V, the accelerating voltage. Hence if high sensitivity is desired, V, should be small. This, however, conflicts with another requirement, i.e. high energy of the electrons where they strike the fluorescent screen or recording surface. These conflicting requirements have been met in certain tubes by means of an additional electrode known as an intensifier or post-accelerating electrode. This arrangement is shown by Fig. 17-15, where the electrode takes the form of a conducting ring on the inner surface of the tube close to the screen. Its purpose is to impart to the electrons a further acceleration after their deflexion by the deflecting plates. Thus increased sensitivity may be attained by a relatively low value of V and the energy of the beam increased after deflexion by a post-accelerating potential V_1. The value of the latter is usually equal to that of V with the result that the beam energy is equal to that of an ordinary tube, but the sensitivity is twice as great. In order to avoid axial acceleration in the vicinity of the deflecting plates, the anode is connected to a further electrode in the manner shown.

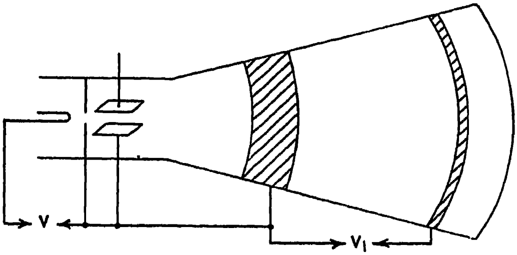


FIG. 17-15

The foregoing theory of electrostatic deflexion is based on the assumptions of a uniform field and an absence of space charges between the plates. With regard to the former a fringe field, of course, exists at the plate edges and the importance of this in modifying the result increases with an increase in the ratio d/l. Formulae of the form

$$l_e = l + Kd$$

have been given, where l_e is the effective plate length and K a constant. In practice the discrepancy between measured and calculated deflexions may amount to as much as 30 per cent. However, one seldom depends on calculated values of the sensitivity but derives this from measured deflexions with known steady potentials.

MAGNETIC DEFLEXION

In order to produce a magnetic deflexion, a magnetic field is created at right-angles to the beam, usually by means of a pair of similarly wound coils attached to the neck of the oscillograph tube. Referring to Fig. 17–16, the field is assumed to be contained in a circular space and to be perpendicular to the plane of the paper.

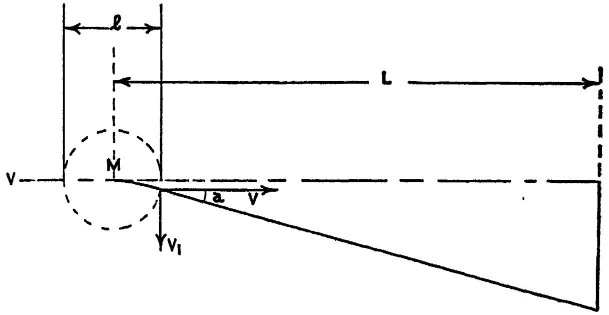


FIG. 17–16

As shown in Chapter III, an electron entering the field experiences a force, perpendicular to its direction of motion, equal to Hev. The acceleration imparted to the electron is

$$\alpha = Hev/m \quad . \quad . \quad . \quad (17\text{–}7)$$

At a time t after entering the field, the axial distance travelled by an electron is

$$x = vt \quad . \quad . \quad . \quad . \quad (17\text{–}8)$$

In the same interval the electron acquires a transverse velocity component equal to

$$v_1 = \alpha t = Hex/m \quad . \quad . \quad . \quad (17\text{–}9)$$

and travels a vertical distance $\frac{1}{2}\alpha t^2$ which, from (17–7) and (17–8), is equal to

$$\tfrac{1}{2}He/m \,.\, x^2/v \quad . \quad . \quad . \quad (17\text{–}10)$$

Thus the trajectory of the electrons (and the beam) within the field is a parabola.

From (17–8) the transit time of an electron in the field is

$$t = l/v$$

and the vertical velocity component acquired in this time is

$$v_1 = Hel/m \quad . \quad . \quad . \quad (17\text{–}11)$$

Referring to Fig. 17–16, the exit velocity of an electron is given by tan $a = v_1/v$ which, from (17–11), is equal to

$$Hel/mv$$

But $v^2 = 2eV/m$ and, therefore,

$$\tan a = Hlv/2V$$

From (17–10), the vertical displacement is

$$\tfrac{1}{2} \,.\, He/m \,.\, l^2/v$$

$$= \frac{Hlv}{2V} \cdot \frac{leV}{mv^2} = \frac{l}{2} \cdot \frac{Hlv}{2V} = \frac{l}{2} \,.\, \tan a$$

Thus, as with electrostatic deflexion, the base of the triangle of Fig. 17–16 is equal to $l/2$, i.e. half the length of the field. Similarly, the electron path and beam may be regarded as linear and originating at a virtual cathode, M, midway between the edges of the field. Hence the deflexion of the beam at the recording surface is

$$\begin{aligned} D &= L \tan a \\ &= LHlv/2V \\ &= lLH\sqrt{e/2mV} \end{aligned}$$

The sensitivity in this case is referred to unit field strength and hence is

$$lL\sqrt{e/2mV}$$

or

$$0{\cdot}3lL/\sqrt{V} \text{ cm./gauss} \quad . \quad . \quad . \quad (17\text{–}12)$$

As in the electrostatic case, practical results are likely to differ from those indicated by (17–12). This is because of the fringing effect of the field and the difficulty of maintaining a uniform field over the whole operating length. As the sensitivity is proportional to L, the coils are placed as far from the recording surface as is practicable. The deflexion factor is, of course, given by

$$\sqrt{V}/0{\cdot}3lL \text{ cm./gauss}$$

EFFECT OF ELECTRON TRANSIT TIME ON SENSITIVITY

The formulae of (17–6) and (17–12) have been developed on the assumption that the transit time of the electrons between the plates is small compared with the periodic time of the wave undergoing examination. Thus these formulae represent what may be termed the static sensitivity of a tube. If the transit time is of the same order as the periodic time, errors occur because the sensitivity is no longer given by (17–6) and (17–12). To consider the conditions under these circumstances let a sinusoidal voltage, represented by $V_p = V_{pm} \sin pt$, be applied to the deflecting plates, where $p = 2\pi f$, f being the frequency. If the voltage is $V_{pm} \sin pt_1$ at the moment an electron enters the plate space, it will be $V_{pm} \sin p(t_1 + t)$ when it leaves, t being the electron transit time equal to l/v. Hence the accelerations at the entrance and exit of the plates are, respectively,

$$\frac{eV_{pm}}{dm} \sin pt_1$$

and

$$\frac{eV_{pm}}{dm} \sin p(t_1 + t)$$

and at time $(t_1 + t)$

$$\frac{dv_1}{dt} = \frac{eV_{pm}}{dm} \sin p(t_1 + t)$$

Integrating

$$v_1 = \frac{eV_{pm}}{pdm} [\cos p(t_1 + t)]_t^0$$

$$= \frac{eV_{pm}}{pdm} [\cos p(t_1 + t) - \cos pt_1]$$

The tangent of the angle (see page 553) at which the electron leaves the field is

$$\tan \alpha_1 = \frac{v_1}{v} = \frac{eV_{pm}}{pvdm} [\cos p(t_1 + t) - \cos pt_1]$$

Now in the static case, $\tan \alpha = leV_{pm} \sin pt_1/mdv^2$ and hence

$$\frac{\tan \alpha_1}{\tan \alpha} = \frac{1}{pt \sin pt_1} [\cos p(t_1 + t) - \cos pt_1]$$

But according to page 553 the deflexion of an electron is proportional to $\tan \alpha$ and thus

$$D_d = \frac{D_s}{pt \sin pt_1} [\cos p(t_1 + t) - \cos pt_1]$$

$$= \frac{D_s}{pt \sin pt_1} 2 \sin p(t_1 + t/2) \sin pt/2$$

$$= D_s \frac{\sin (pt_1 + pt/2)}{\sin pt_1} \cdot \frac{\sin pt/2}{pt/2} \qquad (17–13)$$

where D_d and D_s are, respectively, the dynamic and static deflexions for the same peak voltage V_{pm}. If we now take the ratio of the dynamic and static sensitivities as being proportional to the ratio of the respective deflexions under the same peak voltage then

$$S_d = S_s \frac{\sin (pt_1 + pt/2)}{\sin pt_1} \cdot \frac{\sin pt/2}{pt/2} \quad . \quad . \quad (17\text{–}14)$$

where S_d and S_s are, respectively, the dynamic and static sensitivities. This result at once shows that S_d only equals S_s when p, i.e. $2\pi f$, is very small. It is also evident that the dynamic sensitivity depends on t_1, i.e. on the value of V_p at the moment the electron enters the plate field. This is, perhaps, not altogether unexpected for the change in V_p during the transit time depends on the rate of change of $V_{pm} \sin pt$, i.e. on $pV_{pm} \cos pt$. Considering the peak deflexion, this occurs for $pt_1 = \pi/2$ and

$$D_d = D_s \cos pt/2 \frac{\sin pt/2}{pt/2}$$

$$= D_s \frac{\sin pt}{pt} \quad . \quad . \quad . \quad . \quad (17\text{–}15)$$

Thus, the deflexion will have peak values for $pt = n\pi/2$, each of which will be inversely proportional to the frequency of the deflecting voltage. For $pt = n\pi$ the deflexion will be zero, n being any integer. If we replace t in (17–13), (17–14), and (17–15) by l/v, it is evident that D_d and S_d in addition to depending on frequency also depend on the plate length and the velocity of the electron. As the latter varies as $\sqrt{V}$, D_d and S_d are, therefore, functions of the anode accelerating voltage. Hence for high-frequency work the plate length should be short and the anode voltage high. This, it will be noted, is the reverse of the conditions desirable for large deflexions at low frequencies.

Time Bases

When the beam of a cathode-ray tube is subjected to the influence of a single electric or magnetic field only, the result is the expansion of the fluorescent spot into a line. The length of the line is a measure of the magnitude of the electric or magnetic field causing the expansion, but, beyond this, yields little other information. The general purpose of a cathode-ray tube is, of course, to indicate the variations of one quantity with respect to another, and to effect this, composite deflexion is necessary by two or more simultaneous deflexions. The resulting image on the screen or photographic plate is, of course, two-dimensional, and may be required in either cartesian or polar co-ordinates, generally the former. Where cartesian co-ordinates are required, the deflexions are arranged to occur along mutually perpendicular axes. Naturally the deflexions must be quite independent.

The most commonly employed deflexion is double-electrostatic. In these circumstances the fields may be superimposed by the arrangement of Fig. 17–17 or staggered by means of the plate

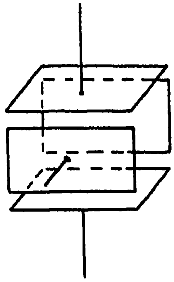


FIG. 17–17

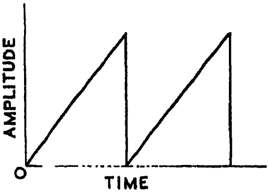


FIG. 17–18

system of Fig. 17–5. Fig. 17–17 does not secure complete independence of the fields and this results in the arrangement of Fig. 17–5 being commonly employed. With the latter the distance of adjoining edges of the plates must be greater than the extent of the field fringes if independence is to be secured. The distance between the plates nearer the anode may be smaller than that between the further pair if desired on account of the smaller movement of the beam at the former. From (17–6) it is apparent that the sensitivities of the two sets of plates may be different unless the plate dimensions are such as to avoid this.

As stated above, a cathode-ray tube is employed to examine the variation of one quantity with respect to another, this resulting in a two-dimensional image. If the result is desired in rectangular co-ordinates, then it is clear that the independently variable quantity must form the horizontal or x-ordinate and the dependent variable the vertical or y-ordinate. In a great many instances it is, of course, desired to trace some phenomenon against a base of time. In the majority of instances this necessitates an X-plate deflexion which increases linearly with time. If an image is required for visual examination (rather than for photographic recording), then a single excursion of the spot must undergo rapid repetition over the same path in order that the persistence of vision may create the illusion of a continuous figure. This evidently necessitates the periodic return of the fluorescent spot to its starting point, the time of return being negligible compared with that of the forward motion. The desired characteristics, then, of the X-plate deflexion may be represented by Fig. 17–18, in which the amplitude of the deflexion varies directly with time until a predetermined maximum is reached; whereupon the spot returns with great rapidity to its starting point and the cycle is then indefinitely repeated. As the brilliance of any portion of the image varies inversely as the velocity of the fluorescent spot, it is evident that the return of the spot, or "fly-back," may be made at such speed as to be imperceptible. Where the speed is such as not to secure this condition, the return trace may be rendered invisible by "blacking out" the spot by the automatic application of excessive bias to the grid of the cathode-ray tube.

In addition to the linear time base outlined above, other time bases exist, such as sinusoidal and circular bases, etc. However, the linear type will first be treated in some detail followed by a description of the characteristics and uses of various other forms.

Linear Time Bases

A description of the necessary action of the linear time base having been given, it is now necessary to consider how this action may be brought about. From Fig. 17–18 it is apparent that the time-base mechanism must generate what may be termed "saw-tooth" oscillations. The relatively slow traverse may be obtained in either one of two ways, viz. (1) by the change in potential across a condenser due to a constant charging current, this being termed a capacitive time base, (2) by the change in current through a pure inductance due to the application of a constant voltage (inductive time base). Both forms require some switching arrangement in order, at some predetermined moment, to change rapidly the traversing action to the fly-back condition. The switching may be effected by the time base itself or by some external agent. Circuits will first be considered in which the switching is self-operated by means of an arrangement of valves. The valves may be of either the soft or hard types and we shall now proceed to a study of linear capacitive time bases of the soft-valve type.

Probably the earliest form of linear time base is that indicated by Fig. 17–19. Fundamentally this consists of a non-inductive resistance R, a capacitance C, and a neon lamp N, arranged in the manner shown. The lamp is of the two-electrode, cold-cathode, type, and is filled with neon at a low pressure with a slight trace of argon or hydrogen added. Such lamps are exemplified by the Osglim beehive or honeycomb lamp. These lamps have a striking, or ignition, potential of about 170 volts and an extinction potential of about 130 volts. At the moment of closing the switch S, the circuit conditions are given by

$$iR + v = E$$

where i and v are, respectively, the instantaneous values of the condenser current and voltage. Now $i = Cdv/dt$ and hence

$$CR\frac{dv}{dt} + v = E$$

from which

$$\frac{dv}{E - v} = \frac{dt}{CR}$$

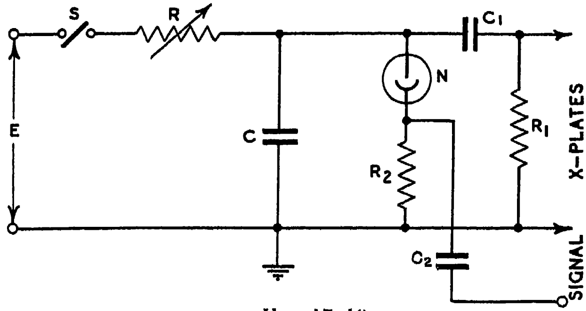


FIG. 17-19

Integrating $\log_e K(E - v) = -t/CR$

and $K(E - v) = \varepsilon^{-t/CR}$

where K is a constant. When $t = 0$, $v = 0$, and thus $K = 1/E$. Hence

$$v = E(1 - \varepsilon^{-t/CR}) \quad . \quad . \quad . \quad (17\text{–}16)$$

Now when v reaches the striking potential of the neon lamp, a discharge will occur, the condenser voltage rapidly falling to v_e, the lamp extinction potential. At extinction, the condenser no longer being shunted, the voltage will again rise to the lamp striking potential v_s and the cycle of voltage variation, i.e. v_e to v_s, will be indefinitely repeated. If t_1 is the time for the condenser voltage to rise from v_e to v_s, then

$$t_1 = CR \log_e \frac{E - v_e}{E - v_s}$$

and assuming the discharge time to be negligible compared with t_1, the oscillation frequency is

$$f = \frac{1}{CR \log_\varepsilon \dfrac{E - v_e}{E - v_s}} \quad . \quad . \quad . \quad (17\text{–}17)$$

The foregoing analysis shows that the circuit of Fig. 17–19 is capable of generating relaxation oscillations having a frequency given by (17–17) and an amplitude equal to $(v_s - v_e)$ volts. The voltage on C at any instant after the initial curve, given by (17–16), has been traced may be found from the initial conditions $v = v_e$, when $t = 0$. Thus

$$v = (E - v_e)\,(1 - \varepsilon^{-t/CR}) + v_e \quad . \quad . \quad (17\text{–}18)$$

Now, although the circuit of Fig. 17–19 has been termed a linear time base, it is evident that the term as applied here is a misnomer for (17–18) is clearly not a linear function of time. What is desired is, of course, that dv/dt should be constant. Differentiating (17–16) and (17–18), we have

$$\frac{dv}{dt} = \frac{E}{CR}\,\varepsilon^{-\frac{t}{CR}} \quad . \quad . \quad . \quad . \quad . \quad (17\text{–}19)$$

$$\frac{dv}{dt} = \frac{E - v_e}{CR}\,\varepsilon^{-\frac{t}{CR}} \quad . \quad . \quad . \quad (17\text{–}20)$$

which show that the slopes of the curves are only constant at $t = 0$. As the velocity with which the spot crosses the screen is proportional to dv/dt the difference in the velocities of the spot at the commencement and the end of its travel is proportional to

$$\frac{E - v_e}{CR}\,(1 - \varepsilon^{-\frac{t_1}{CR}})$$

For given values of t_1 and v_e, it is clear that if approximate linearity is to be achieved, $\varepsilon^{-t_1/CR}$ must be as near unity as possible. This means that CR should be large, also E.

In order to apply the voltage oscillation on C to the X-plates of the tube, the combination C_1R_1 is employed. The purpose of this is to ensure that when the spot is expanded by the time base to a line, the centre point of the latter coincides with the centre of the screen. In the absence of C_1 the mean d.c. potential of C, i.e. $(v_s + v_e)/2$, would deflect the beam from the centre of the screen to a given position and a line with an amplitude corresponding to $(v_s - v_e)$ would appear with its centre at this position. Thus, if a Y-plate deflexion were produced, the resulting image would be non-central. With C_1 present, the d.c. potential is absorbed by this, only the a.c. component occurring across R_1. As the amplitudes of the half-waves of the a.c. component are each equal to $(v_s - v_e)/2$, the mean position of the spot is the centre of the screen with equal deflexions, corresponding to $(v_s - v_e)/2$, to left and right of this. Hence with a Y-plate deflexion a centrally placed image results.

The effect of two simultaneous deflexions is, generally, to cause the fluorescent spot to describe a curved path. Providing the deflecting frequencies are sufficiently high, persistence of vision causes a curve to appear on the fluorescent screen. For visual observation it is, of course, desirable that the curve shall appear motionless, a curve in this condition being termed a stationary or "standing" figure. To ensure this, the frequency of the time base must be a sub-multiple of the frequency of the phenomenon under observation. In these circumstances an integral number of waves will appear on the screen and a standing figure will result. If the above frequency condition is not realized, then the result may vary from a slow motion of the image either backwards or forwards across the screen to complete confusion. It follows that control of the time-base frequency is essential to produce standing figures. In practice this is achieved by variation of R and C. R is continuously variable while a bank of condensers is employed, the appropriate condenser being selected by means of a selector switch.

In Fig. 17–19 there will be noted a resistance-condenser combination R_2C_2. The purpose of this is to secure exact synchronism of the signal and time-base frequencies in order to "lock" the image. Without this feature, it is extremely difficult to obtain a perfectly stationary image by variation of the product RC. The subject of synchronization will be dealt with at greater length after further time-base circuits have been considered.

The lack of linearity of the circuit of Fig. 17–19 may be overcome to some extent by replacing the resistance R by a saturated diode. In this case the condenser charging current tends to be constant, resulting in the voltage on C increasing linearly with time. Because of the lack of true saturation with valves having coated cathodes, a valve with a plain tungsten filament should be employed. The condenser charging current and time-base frequency are varied by variation of the filament temperature. As an alternative to, and an improvement on, this arrangement, a pentode, operating above the knee of the anode-voltage/anode-current characteristic, may be employed. In this case the time-base frequency is controlled by variation of either the screen or grid potential. In general, with both the above methods, the value of E required is less than with a resistance condenser combination, but care must be taken to ensure that the valve anode voltage does not pass into the curved portion of the anode-voltage/anode-current characteristic during the charging of the condenser. If t is the condenser charging time and I the constant charging current, then,

$$It = C(v_s - v_e)$$

from which
$$t = \frac{C(v_s - v_e)}{I}$$

and
$$f = \frac{I}{C(v_s - v_e)}$$

where f is the time-base frequency and C is expressed in farads.

THE THYRATRON TIME BASE

The voltage efficiency of a time base is expressed as

$$\frac{\text{Voltage Excursion of Condenser}}{\text{Voltage of d.c. Supply}}$$

Because of the low value of $(v_s - v_e)$ in Fig. 17–19, it is clear that this time base has a low voltage efficiency. A further disadvantage is the small scan produced. The value of $(v_s - v_e)$ is about 40 volts and with a tube sensitivity of 1 mm. per volt the length of the scan is only 4 cm. Both of these shortcomings are largely removed by replacing the neon lamp of Fig. 17–19 by a thyratron, preferably of the gas-filled type. The method of employing a thyratron for time-base work is indicated by Fig. 17–20. By varying the value of the bias, the striking potential may be varied and with it the amplitude of the scan or sweep in the cathode-ray tube. It will be appreciated that a much greater scan is possible with a thyratron than with a neon lamp. For example, with a bias of 4 volts, a control ratio of 50 and an ionizing potential of 20 volts, the voltage variation on the condenser is $[(4 \times 50) - 20] = 180$ volts, which applied to a tube having a sensitivity of 1 mm. per volt gives a scan of 16 cm. Of course, the maximum scan obtainable with a thyratron is that corresponding to the maximum anode-cathode potential that the thyratron will withstand.

A further advantage of the thyratron time base is its method of synchronization. As indicated by Fig. 17–20 this is effected by applying a portion of the work voltage to the grid in the manner shown. This, of course, has the effect of varying the striking potential of the tube. Because of the large control ratio, it follows that synchronization may be effected with a far smaller voltage than in the case of the neon lamp and with a consequently smaller loading on the work voltage. It will be noted that a resistance is included in the anode lead. The purpose of this is to limit the discharge current through the valve which, if allowed to exceed a certain value, has a detrimental effect upon the life of the valve. The minimum value of this resistance is given by E/I, where E is the maximum scan voltage and I the instantaneous peak current rating of the valve. The value of this resistance should not be higher than is absolutely necessary, for it retards the rate of discharge of the condenser, thus reducing the velocity of the fly-back. As an alternative method of limiting the discharge current, a separate inductance may be connected in series with each of the various condensers in the time-base bank. The maximum energy stored in a condenser is $\frac{1}{2}CE^2$ and, if I is again the peak current, then

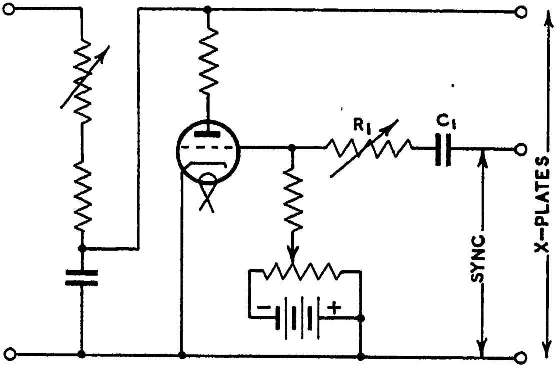


FIG. 17–20

$$\tfrac{1}{2}LI^2 = \tfrac{1}{2}CE^2$$

and

$$L = \frac{CE^2}{I^2}$$

In addition to the anode resistance it will be seen that a resistance is also included in the grid circuit. This, as explained in Chapter IX, is to limit the grid current during the ionizing period. The value of this resistance should not be too high, for a high grid resistance tends to render the grid potential susceptible to pick-up from other parts of the circuit, such as the mains supply. etc. The purpose of the combination C_1R_1 is for the application of a portion of the work voltage for synchronization. R_1 is large to avoid unduly loading the work voltage, while C_1 isolates the grid from the d.c. component in this voltage.

A modification of the thyratron time base, due to Blumlein, is shown by Fig. 17–21. It will be noted that inductances are included in both the charging and discharging circuits. That in the latter is relatively small, while L is of the order of 30 henrys. The circuit conditions during charging are expressed by

$$L\frac{di}{dt} + Ri + v = E$$

or

$$\frac{d^2v}{dt^2} + \frac{R}{L}\frac{dv}{dt} + \frac{v}{LC} = \frac{E}{LC}$$

where v and i are, respectively, the instantaneous values of the condenser voltage and current. The general solution of the above equation is

$$v = \varepsilon^{-\frac{Rt}{2L}}\left(A\sin\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}\,t + B\cos\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}\,t\right) + E$$

and, taking the initial conditions as $t = 0$, $v = -v_1$, and $dv/dt = 0$, we have

$$v = E - \varepsilon^{-\alpha t}(v_1 + E)\left(\frac{\alpha}{w}\sin wt + \cos wt\right)$$

or

$$v = E - \varepsilon^{-\alpha t}(v_1 + E)\sqrt{1 + \left(\frac{\alpha}{w}\right)^2}\sin(wt + \phi)$$

where

$$\alpha = R/2L,\ w = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}},\ \phi = \text{arc tan}\,\frac{w}{\alpha}$$

In order to simplify the understanding of this circuit we shall assume R to be negligible and v_1 to be equal to E. Then

$$v = E(1 - 2\cos wt)$$

and, if the charging time is such that $wt = \pi/2$, the condenser voltage changes from $-E$ to $+E$, i.e. a total change of $2E$ volts. The simplified expression for v is not, of course, a linear function of time, but in practice the effect of R is to have a straightening effect on the characteristic. It also reduces the total change in condenser voltage. However, the point which it is desired to make is that a much greater scan is possible with the circuit of Fig. 17·21 than when the condenser is charged through a pure resistance.

It remains to be considered why the initial condition was taken of $t = 0$, $v = -v_1$. At the moment the condenser discharges through the thyratron, the voltage on the latter drops to the ionizing potential of the gas-filling of this. The discharge circuit is then the essential oscillatory one of L_1C with, of course, the proviso that, because of the valve, the current cannot reverse. The current, then, is interrupted when about to make its first reversal, i.e. half a period after the commencement of the condenser discharge. But if the condenser voltage were $+v_1$ at the commencement of the discharge, it must be $-v_1$ half a period later. The periodic time of the discharge is approximately $2\pi\sqrt{L_1C}$ and hence the discharge period is $\pi\sqrt{L_1C}$ seconds.

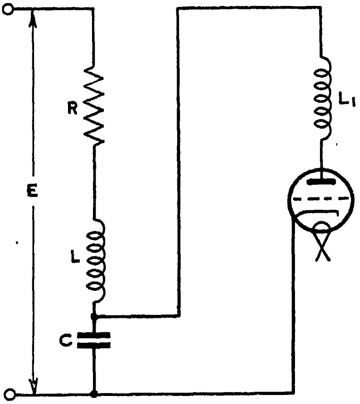


FIG. 17–21

HARD-VALVE HIGH-FREQUENCY TIME BASES

The foregoing time bases are of the soft-valve type and, providing the desired frequency range is not too high, that of Fig. 17–20 in particular may be relied upon to give excellent results. It will be noted that due to the small discharging time it is only necessary to switch the discharge circuit, the charging circuit actually functioning continuously. However, if such time bases are operated at high frequencies, the charging time becomes comparable with the discharging time and ultimately the time base ceases to function. This is because the charging rate becomes so rapid that sufficient time is not available for deionization to take place and, in consequence, the supply current passes continuously through the thyratron. In general, it may be said that the upper frequency limit of a time base employing a thyratron is of the order of 10^5 cycles per second.

Hard-valve time bases may be classified under two headings: (*a*) those which are self-operated, usually taking the form of a modified valve oscillator; (*b*) those which are discharged by an independent voltage control. Considering those under (*a*), these are similar to a thyratron base in that they consist of a condenser-charging device and a discharging or triggering device. In many circuits the condenser charging is effected in an exactly similar manner to that of a soft-valve time base, i.e. by means of a non-inductive resistance, saturated diode, or pentode. The principal point of departure is in the triggering circuit employed for discharging the condenser. The main difficulty in the design of hard-valve time bases is the imitation of the abrupt action of the thyratron discharge. It might appear that a thyratron could be replaced by a hard valve biased to cut-off at an anode potential equal to that at which it is desired to discharge the condenser. However, when it is remembered that ionization is negligible in hard valves and that in consequence the grid has continuous control, it is evident that this is impossible unless by some means the grid bias may be suddenly reduced and, preferably, driven positive at the desired moment. This last proviso contains the method by which many hard-valve time bases operate, the change in grid condition being brought about either by the time base itself (self-operated) or by some external agency.

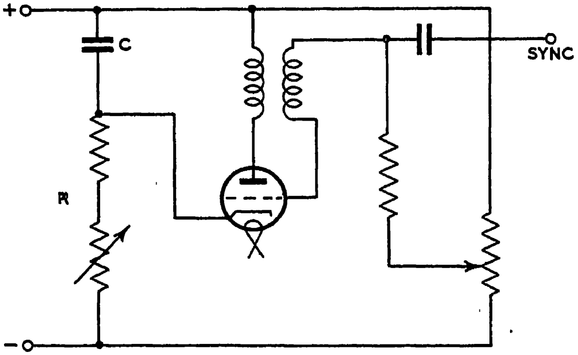


Fig. 17–22

As an example of a self-operated time base working on the foregoing principal, Fig. 17–22 may be considered. Across the condenser (which may be charged by any one of the methods previously described) is connected a triode with the primary of a transformer in the anode lead. The transformer secondary is connected in the grid circuit in such a manner that a discharge from the condenser through the triode reduces the steady bias and drives the grid positive. Considering now the circuit from the moment that the voltage on C is zero, the condenser will commence to charge through the resistance R. At a voltage depending on the value of the steady bias, anode current will commence to flow, the e.m.f. induced in the secondary will drive the grid positive, causing the anode current to increase cumulatively and rapidly discharge the condenser. The e.m.f. induced in the transformer secondary winding is Mdi/dt, where M is the mutual inductance between the windings and di/dt the rate of change of anode current. As the condenser becomes discharged, di/dt will ultimately decrease and become zero. At this moment the discharge current is a maximum and the steady bias alone affects the grid. Beyond this point the anode current commences to decrease, di/dt becomes negative, the grid very rapidly regains control, and the discharge ceases. The cycle of events is then repeated indefinitely. It is clear that the amplitude of the sweep with this time base depends on the value of the steady bias and may be conveniently adjusted by varying this. If automatic grid bias is employed, the bias resistance smoothing condenser should be large in order that a steady bias may be developed. The value of this bias is proportional to the average anode current and tends to stabilize the time-base frequency. Thus, if the supply voltage should increase and in consequence the time-base frequency, the bias and scan will also increase, this tending to counteract the frequency increase.

The time base may be synchronized by a third transformer winding, upon which a portion of the work voltage is impressed. Alternatively, a resistance may be included in the grid circuit in the manner shown.

The foregoing description of the operation of this time base has ignored the possible effects of the self-capacitances of the transformer windings. Because of these, the circuit may act as a high-frequency oscillator, the grid being swung alternatively positive and negative, and the condenser discharged in a series of impulses. Actually it is found that the circuit tends to give better results when operated by means of a low or audio-frequency transformer, because then the discharge occurs for one cycle of grid voltage only. This, of course, reduces the frequency range of the time base, with the result that its application is chiefly to low-frequency work, say, up to 20 kilocycles. In order to obtain satisfactory operation it is important that the transformer should be damped, this being effected by connecting a non-inductive resistance in parallel with the primary winding. The optimum value of this resistance must be determined experimentally.

The time base in its low-frequency form finds application in television circuits, such as producing the scanning frequency.

A high-frequency time base, due to Puckle, is shown by Fig. 17·23. In this circuit the condenser is charged by means of a pentode and discharged by a pair of hard valves which, in combination, imitate the action of a thyratron. During charging, the triode is biased to cut-off by the volt drop on the anode resistance of the discharge pentode, V_3. Under these conditions there is no current in R_1 and, consequently, V_3 is free of bias and the current in R_2 is a maximum. This, of course, results in the bias on the grid of V_2 also being a maximum. Thus, it is quite possible with a supply voltage of, say, 500 volts to have the grid of V_2 400 volts or so negative with respect to the positive supply line. As the condenser C becomes charged, the cathode of V_2 will run negative and the voltage on C will ultimately reach a point at which current will commence to flow in V_2 and R_1. The effect of the volt drop on R_1 will now send the suppressor grid of V_3 negative, resulting in a reduction in current through R_2. The change in potential on R_2 reduces the bias on the grid, drives it positive, increasing the current through V_2. It will be readily appreciated that these actions between V_2 and V_3 are cumulative, leading to a rapid discharge of C. With C discharged, current in R_1 ceases, bias is removed from V_3, the current regains its maximum value in R_2 and full bias is again applied to V_2. Hence the cycle of events is indefinitely repeated.

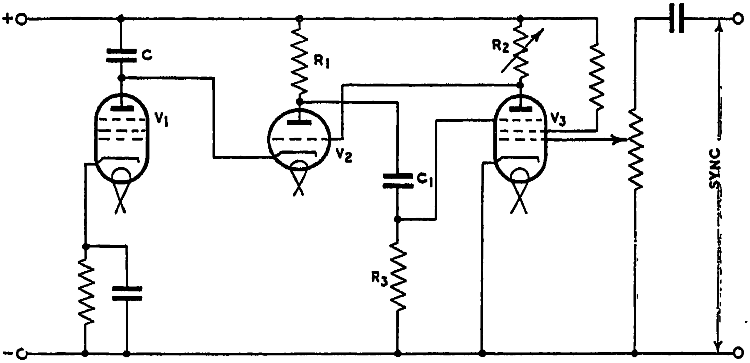


FIG. 17–23

Actually the rate at which C can be discharged is limited by the fact that when the grid of the triode goes positive grid current from this valve flows in the anode load of V_3. Thus the rate at which the anode of V_3 (and the grid of V_2) can go positive is limited. As the grid of V_2 is going positive so also is the cathode of this valve, and it appears in practice that they both travel positive at approximately the same rates.

It is evident that the value of R_2 controls the moment at which C discharges. Hence the amplitude of the scan may be varied by varying R_2. The effect of varying R_1 is to affect the fly-back velocity. If R_1 is increased, then the fly-back velocity is decreased. However, the greater the value of R_1 the greater is the rapidity with which the change-over from the charging to discharging state occurs. Some values given by Puckle for various components of the circuit are—

$$R_1 = 2000 \text{ ohms (max.)}$$
$$R_2 = 0{\cdot}25 \text{ megohm}$$
$$C_1 = 0{\cdot}1\ \mu\text{F}$$
$$R_3 = 1{\cdot}0 \text{ megohm}$$

Synchronization is effected in the manner indicated by Fig. 17–23. A portion of the work voltage is fed to the control grid of V_3. This is amplified by V_3 and causes the potential of the grid of V_2 to fluctuate above and below that value determined by R_2, thus varying the condenser discharge potential.

It may be mentioned that voltage impulses of approximately square wave-form may be derived from either R_1 or R_2. If desired, these may be employed for blacking-out the fly-back on the cathode-ray tube. This is effected by connecting the anode of V_2 to the grid of the tube by means of a condenser.

The foregoing time base will operate at frequencies up to 10^6 cycles per second.

THE TRANSITRON TIME BASE

A single-valve time-base circuit, operating on the principle of the transitron oscillator, is shown by Fig. 17–24. Assuming the condenser C_1 to be discharging through the pentode, the voltage on the latter will continually fall, the discharge current at first remaining approximately constant. As the knee of the anode-voltage/anode-current characteristic is approached, the anode current will commence to decrease. As, however, the total cathode current tends to remain constant the screen current will increase, causing the screen voltage to fall. This causes the condenser C_2 to discharge through R_3 and the valve, with the result that the upper end of R_3 becomes negative relative to the cathode, this, of course, driving the suppressor negative. The effect of the reduced screen voltage and the negative suppressor reduces the anode current still further, and the process becomes rapidly cumulative, resulting in a cut-off of the anode current. The condenser C_1 now commences to recharge through R_1, the negative potential of the suppressor slowly decreasing as C_2 discharges. At some voltage on C_1 current will commence to flow to the anode, which current will be withdrawn from the screen circuit. This causes the screen potential to rise and a charging current to flow from the h.t. supply through R_3 into C_2. This reverses the direction of current flow in R_3 and, in consequence, the suppressor potential rapidly changes from negative to positive. The rise in screen potential and the reversed suppressor potential now cumulatively increase the anode current, with the result that C_1 is again rapidly discharged.

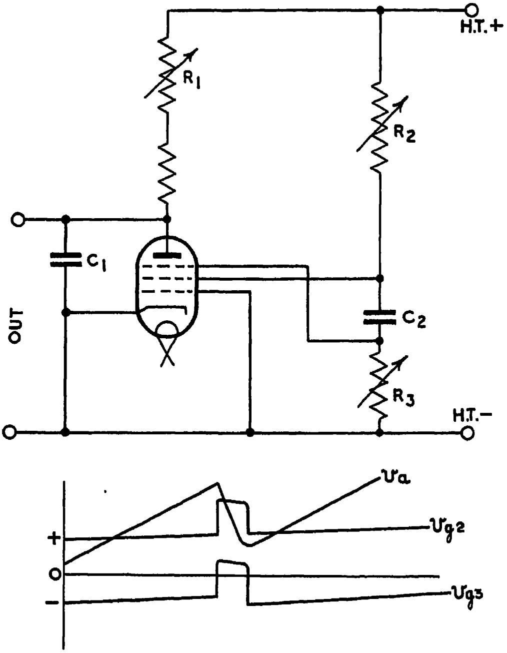


FIG. 17–24

The various waveforms during the circuit operation are shown by Fig. 17–24. The rate of change of voltage on C_2 depends, of course, on the time constant C_2R_3. However, while C_2 is discharging, the valve is included in the circuit with the result that the rate of change of potential on R_3 is less than when C_2 is charging. It is evident that in order to hold the suppressor negative during the charging stroke of C_1, C_2 should not be too small. Synchronization may be effected by impressing a portion of the work voltage on the grid. For the purpose of considering the characteristics of the circuit the following components are suggested—

$R_1 = 0{\cdot}1$ to $2{\cdot}0$ megohms $\quad C_1 = 0{\cdot}1\ \mu F$

$R_2 = 0{\cdot}05$,, $0{\cdot}25$ megohm $\quad C_2 = 1{\cdot}0\ \mu F$

$R_3 = 0{\cdot}01$,, $0{\cdot}25$,, $\quad V =$ Mullard SP4

H.T. 500 volts

Synchronization

Reference has already been made to the subject of synchronization of the time-base and work frequencies. Synchronization is effected by the application of the work voltage to the time base in such a manner that it initiates either the scan or the fly-back. In the case of a capacitative time base it is the latter which is initiated. Referring to Fig. 17–19, it will be noted that the work voltage, or a portion of it, is applied to a resistance in series with the neon lamp. This means that a portion of the work voltage is superimposed on the time-base voltage and thus the moment of firing of the neon lamp may be varied by the work voltage itself. The manner in which this is effected will be clear from Fig. 17–25, in which the time-base voltage is represented by Curve 1 and the superimposing work voltage by Curve 2. Assume that synchronism has been effected, the lamp firing at P_0 with a striking potential of 150 volts. With a 50-cycle per second sinusoidal wave applied to the oscillograph Y-plates, one complete wave now appears on the screen, the time of the sweep being 20 milliseconds. Now although small

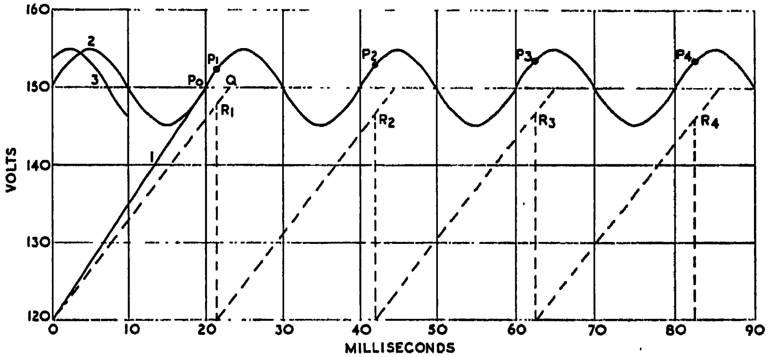


FIG. 17–25

changes in time-base and work frequencies may occur, or changes in the striking and extinction potentials of the lamp, synchronism will be automatically retained, as will now be shown.

Let it be supposed that for some reason the time-base frequency falls, the broken line OQ now representing the time-base voltage. However, due to the superimposition of the work voltage, the lamp fires at R_1 instead of Q, because at this point the sum of the time base and superimposed work voltages are equal to the lamp-striking voltage. Thus the scan time is less than it would be if striking occurred at Q and for the case shown is approximately 21 milliseconds. The next oscillogram commences at P_1 and terminates at P_2 after a period of 20·7 milliseconds. The following oscillogram commences at P_2 and terminates at P_3 after approximately 20 milliseconds. But this period is equal to that of the work voltage and thus, at this stage, synchronization again results. It will be readily appreciated that with each succeeding cycle firing occurs at a later moment on the work-voltage curve. This permits a larger portion of the work voltage to be added to that of the time base to cause striking, thus permitting a reduced scan to compensate for a reduced rate of rise of condenser voltage. The appearance of the wave on the screen is now that shown between P_3 and P_4 instead of $150 - P_0$. Further, it may be noted that the scan voltage has dropped from 150 to 147 volts, resulting in a shortening of the picture by $3/30 = 10$ per cent. This means that the work voltage has added 3 volts to that of the condenser in order that striking may occur after a 20-millisecond period. It will be noted that the wave shifts somewhat to the left of its original position, as indicated by Curve 3.

For the case of an increase in time-base frequency we may consider the initial operating conditions to be those of R_4P_4. Then, if the rate of voltage rise of the time-base condenser corresponds to that of OP_0, the conditions will tend to revert to those of P_0 where synchronization will again result. Summarizing, it may be said that any small change in time-base frequency results in an automatic phase displacement of the work and time-base waves resulting in a continuance of the synchronized state. Similar effects occur should the work frequency change, the scan shortening for an increase in frequency and lengthening for a decrease.

The foregoing results may be formulated as follows: The lamp voltage between striking and extinction is $v = v_e + \alpha t$, where t is time and α a constant. At striking

$$v_s = v_e + \alpha t_0$$

where t_0 is the time-base period. If the velocity of the sweep alters, then $v = v_e + \alpha_1 t$ and at time t_0

$$v = v_e + \frac{\alpha_1}{\alpha}(v_s - v_e)$$

The difference between the striking voltage and v is

$$(v_s - v_e)(1 - \alpha_1/\alpha) \qquad . \quad . \quad . \quad (17\text{–}21)$$

Thus, if $\alpha_1 < \alpha$ (slow running), the phase shift is such that a positive voltage, derived from the work, is added to that of the time base.

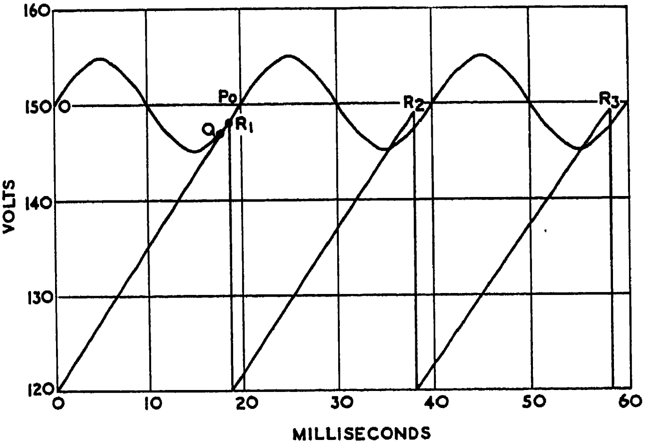


FIG. 17–26

If $\alpha_1 > \alpha$ (fast running), the phase shift is such as to add a negative voltage to that of the time base.

It will be appreciated that changes in the amplitude of the sweep due to variations in v_s or v_e are compensated in a similar manner as changes in work frequency, i.e. an automatic phase shift occurs in order to keep the time-base frequency constant. An example of the effect of a change in v_s is shown by Fig. 17–26. Here it is assumed that the normal value of v_s is 150 volts and that for some reason this drops to 147 volts. Then in the absence of the influence of the work voltage the lamp discharges at Q. However, because of the influence of the work voltage, the discharge occurs at R_1, resulting in an increase in the time-base frequency (19-millisecond period) and a disturbance of the synchronized state. The next discharge occurs at R_2 after a period of approximately 19·5 milliseconds, followed by a third discharge at R_3. By this time the phase shift is such that the time-base period is practically equal to its original value, with the result that synchronization again results.

AMPLITUDE AND PHASE OF SYNCHRONIZING SIGNAL

From what has been stated regarding synchronization, it is evident that the amplitude of the synchronizing signal is a matter of some importance. For example, according to (17–21), the voltage to be introduced by the work to maintain synchronization, should the time-base frequency vary, depends on α_1. Thus the amplitude of the voltage introduced must be at least equal to (17–21) and preferably considerably larger. If less than (17–21), synchronization cannot be effected. As for other reasons it is desirable that the synchronizing signal shall be small, it is necessary to employ means to adjust the signal to the minimum which will give effective locking of the picture. This is usually effected by means of variation of the resistance R_2 in Fig. 17–19. The purpose of the series condensor is to isolate the neon lamp from any d.c. component in the waveform undergoing examination. If the amplitude of the synchronizing voltage is too large, then it may trigger the time base too early during one of the positive half-cycles. This will result in the time-base period no longer being a multiple of that of the work voltage, and synchronization will fail to occur. It, of course, follows that if the time-base period is less than that of the work voltage (time base running too fast) synchronization is again ineffective.

The phase of the synchronizing signal is self-adjusting with respect to the time base as previously demonstrated. Hence the initial phase relationship is of no moment, as automatic adjustment rapidly occurs. When a disturbance of the synchronized state arises it is evident from Figs. 17–25 and 17–26 that automatic adjustment is effective within a few cycles of the disturbance arising and hence is generally quite imperceptible to the eye. The rapidity with which adjustment is effective depends on the steepness of the wave-front of the signal voltage, being greater the steeper the wave front.

SYNCHRONIZATION OF A THYRATRON TIME BASE

As previously stated and indicated by Fig. 17–20, synchronization of thyratron-operated time bases is effected by applying a portion of the work voltage to the grid of the valve. This, of course, has the effect of varying the striking potential of the tube. The method of synchronization in the present case may be appreciated with the assistance of Fig. 17–27. Referring to this, and assuming a steady bias of 6 volts and a control ratio of 25, the firing point of the thyratron is 150 volts. If 1 volt of the work voltage is superimposed on the steady bias, then the firing point of the relay will periodically vary between 175 and 125 volts as shown. Let it be assumed that synchronism has been effected and that the time base then commences to run slow. The time-base voltage will then change from the full line of 1 to the broken line of 2, the firing point changing from P_0 to R_1. It is apparent that at the end of each sweep the relay fires at a lower voltage than that of the preceding one. Finally at R_4 the time-base period is again the same as that of the wave under observation, and synchronization again results.

As in the case of the neon lamp, it is clear that an automatic phase shift occurs in order to lower the striking potential of the relay so that the scan time may be maintained constant. Because of the reduced scan, it is also evident that a slight compression of the picture occurs. Assuming the condenser voltage drops to zero at each discharge, the equation to (1), Fig. 17–27, is $V = \alpha t$ and to (2), $V_1 = \alpha_1 t$. If the scan time for synchronization is t_0

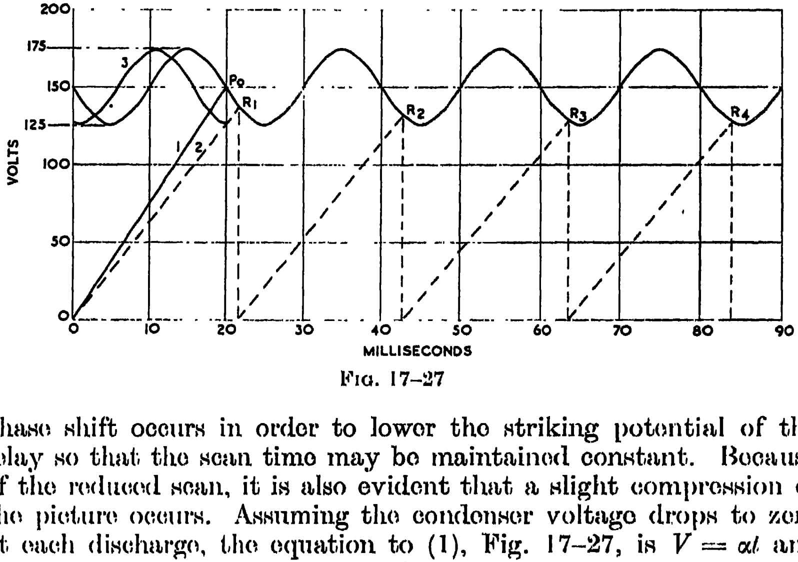


FIG. 17–27

$$V_0 = \alpha t_0$$

$$V_1 = \alpha_1 t_0$$

from which

$$V_0 = \frac{\alpha}{\alpha_1} V_1$$

and

$$(V_1 - V_0) = (\alpha_1 - \alpha)t_0 \quad . \quad . \quad . \quad (17\text{–}22)$$

where V_0 and V_1 respectively correspond to P_0 and R_4. Thus for the case shown by Fig. 17–27 $(V_1 - V_0) = -22$ volts, which indicates that the phase shift must be such as to lower the striking voltage of the relay by this amount. An increase in time-base velocity would mean $\alpha_1 > \alpha$, in which case (17–22) would be positive, and the shift would be such as to raise the striking potential. The cases of a change in work-voltage frequency or of striking potential may be easily worked out with the assistance of a diagram similar to that of Fig. 17–27.

The phase shift due to a slackening of the time-base velocity is shown by (3) in Fig. 17–27. Such shifts are readily demonstrated with a normal oscillograph. For convenience a sine wave is produced on the screen and synchronized. The velocity control is then slowly altered, whereupon the phase shift of the wave may be clearly seen.

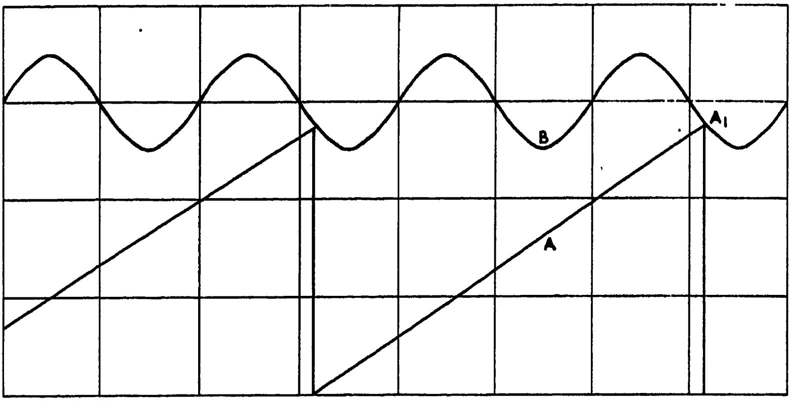


FIG. 17–28

The compression or expansion of the wave is also observable according to whether the time-base velocity is decreased or increased.

Comparing the thyratron time base with that of the neon lamp, it is obvious that, because of the control ratio, a much smaller portion of the work voltage is needed for synchronization in the former case. Thus, for the instance given by Fig. 17·27, the peak alternating voltage needed on the grid is only 22/25 = 0·88 volt, or 0·62 volt, r.m.s. Again, in the former case it is clear that the amplitude of the synchronizing voltage must not be below a certain value to effect synchronization. With the thyratron time base a somewhat interesting phenomenon occurs if the amplitude of the synchronizing signal is continually increased. Assume that a group of waves has been produced on the screen and synchronized as shown by Fig. 17–28. Now, if the synchronizing voltage is continually increased, a moment will occur when the negative peak of the wave *B* makes contact with the time-base voltage curve at *A*. At this moment the synchronizing point will suddenly jump from A_1 to A and one cycle will disappear from the screen. This phenomenon, of course, indicates that the synchronizing signal is too large and is termed "tight" synchronization.

SYNCHRONIZATION OF HARD-VALVE TIME BASES

The synchronization of the time base of Fig. 17–23 is effected in a similar manner to that of a thyratron base. A portion of the work voltage is applied to the grid of V_3 and an amplified and phase-reversed version of this is impressed on the grid of V_2. Thus the "striking" voltage of V_2 will periodically vary in a similar manner to that of a thyratron. The conditions are similar in Fig. 17–22, where a portion of the work voltage fed to the grid of the valve again causes the "striking" voltage of this to vary in the manner of a thyratron.

Circular and Elliptical Time Bases

In certain applications a circular or elliptical time base possesses advantages over those of the linear type. For example, a much longer trace is obtainable, no fly-back is needed, and the trace may be produced by sinusoidal wave forms instead of by the usual saw-tooth. If sinusoidal e.m.f.s are applied to the X- and Y-plates, then the co-ordinates of the fluorescent spot will be given by

$$x = a \cos pt$$

$$y = b \sin pt$$

where a and b are the amplitudes of the e.m.f.s. Squaring and dividing, we have

$$\frac{x^2}{a^2} = \cos^2 pt$$

$$\frac{y^2}{b^2} = \sin^2 pt$$

and
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \sin^2 pt + \cos^2 pt = 1$$

which is the equation to an ellipse. Hence the trajectory of the spot is an ellipse, the elliptical path being traced once for every cycle of the applied e.m.f. If $b = a$, then

$$x^2 + y^2 = a^2$$

and the spot is at a constant distance from the centre of the tube, a circle appearing on the screen. a now constitutes a radius vector and the angle formed by this and the ox axis is given by

$$\tan\theta = \frac{y}{x} = \frac{b \sin pt}{a \cos pt} = \frac{b}{a}\tan pt \quad . \quad . \quad (17\text{–}23)$$

or, as $b = a$,

$$\theta = pt$$

where θ is the angle formed by the radius vector and the ox axis. Hence the radius vector drawn from the spot to the centre of the screen rotates with constant angular velocity.

In order to form either an elliptical or circular trace it is necessary to apply e.m.f.s to the plates which differ in phase by 90°. This may be conveniently done by the phase-splitting device shown by Fig. 17–29. The current is constant in magnitude and phase throughout the resistance-condenser combination and hence the e.m.f.s on R and C are 90° out of phase with each other. The voltage on R is IR and on C, I/wC. Hence for the amplitude of these two voltages to be equal we must have $R = 1/wC$. For this condition, and with equal sensitivities for both pairs of plates, a circular trace will result. If $R \neq 1/wC$ the trace is elliptical. In practice it is desirable to make R variable, as shown, as then different forms of ellipse may be obtained and compensation made for differing sensitivities of the two pairs of plates when a circular trace is desired.

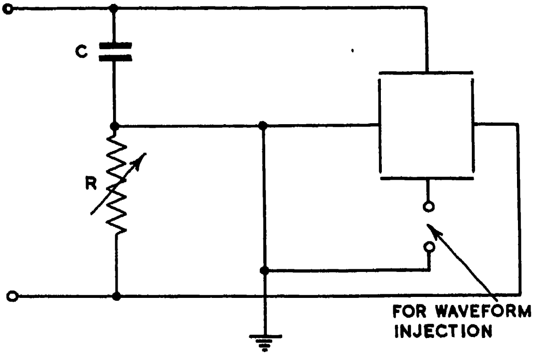


FIG. 17–29

If harmonics are present in the supply source of Fig. 17–29, variations in the velocity and circularity of the time base will occur. This is because the value of R is unaffected by harmonics while $1/wC$ is reduced. These disadvantages may be largely avoided by replacing R by a tuned circuit LC_1 as shown by Fig. 17–30. With this circuit tuned to the supply frequency it is equivalent to a pure resistance of value L/C_1R_1, where R_1 is the resistance of the inductive branch. At harmonic frequencies the impedance of the inductive branch increases, while that of C_1 decreases. Hence the impedances of both C and L/C_1R_1 are reduced by harmonics and the trace is less affected than with the arrangement of Fig. 17–29. With plates of equal sensitivity, and for a circular trace, we must have

$$\frac{L}{C_1R_1} = \frac{1}{wC}$$

or making $C_1 = C$,

$$R_1 = wL$$

But $w^2 = 1/LC$ and, therefore,

$$wL = \frac{1}{wC}$$

and

$$R_1 = \frac{1}{wC}$$

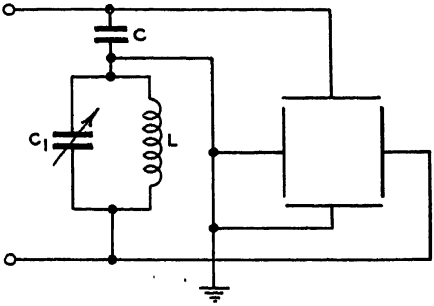


FIG. 17–30

WAVEFORM INJECTION

With the foregoing methods of producing elliptical and circular time bases both pairs of plates are employed, thus leaving no plates available for the waveform to be examined. It is, however, possible to apply the waveform voltage in series with the Y-plates, as shown by Fig. 17–29, in which case the waveform is perpendicularly superimposed on the time-base trace. With a circular trace or a small ellipse this procedure is not very satisfactory, as indicated by Fig. 17–31 (a). Where an elliptical trace is employed, with a wide horizontal sweep, and a major axis which is large compared with the minor axis, better results are obtainable as shown at (b).

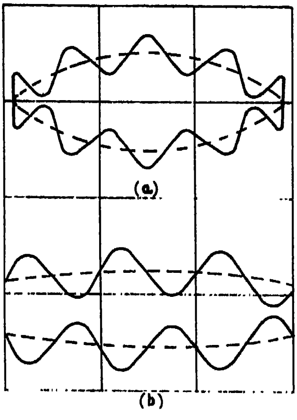


FIG. 17 31

The usual method of applying the work voltage is to place it in series with the anode potential of the tube. In this manner the anode potential is modulated and with it the deflexional sensitivity of the tube in accordance with (17–6). Thus, with a circular trace, a radial motion of the spot is produced and the waveform is spaced round the trace in a symmetrical manner. An effect of the work voltage when applied in this manner is a tendency to defocus the beam. Hence only small deflecting voltages are normally permissible. Larger voltages may be employed if applied to the potentiometer feeding the tube; for in this manner both the anode and grid potentials are simultaneously affected.

A considerably superior, but more complex, circular time-base circuit than those just described is that shown by Fig. 17–32. A phase-splitting circuit is again employed to provide two e.m.f.s which are 90° out of phase. These e.m.f.s are applied to two identical transformers having centre-tapped secondaries. Across the secondaries are four valves arranged for push-pull working, the secondaries providing a four-phase supply to the cathodes and grids of these valves. The work to be examined is simultaneously applied to all four grids and thus modulates the four-phase input and varies the diameter of the circular trace. As the valves are worked in push-pull it may be noted that, in the absence of the phase-splitter supply, the work voltage does not produce a deflexion of the spot. The purpose of the condensers C_1, C_2, C_3, C_4 is to provide a low impedance for the work voltage should this be of high frequency.

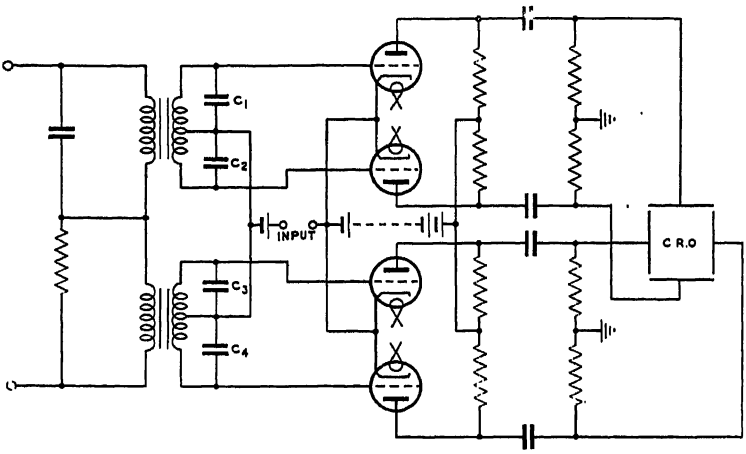


FIG. 17–32

A particular use of a circular time base is the comparison or measurement of frequencies. The time-base frequency is adjusted until a standing figure is obtained, when the frequency of the work

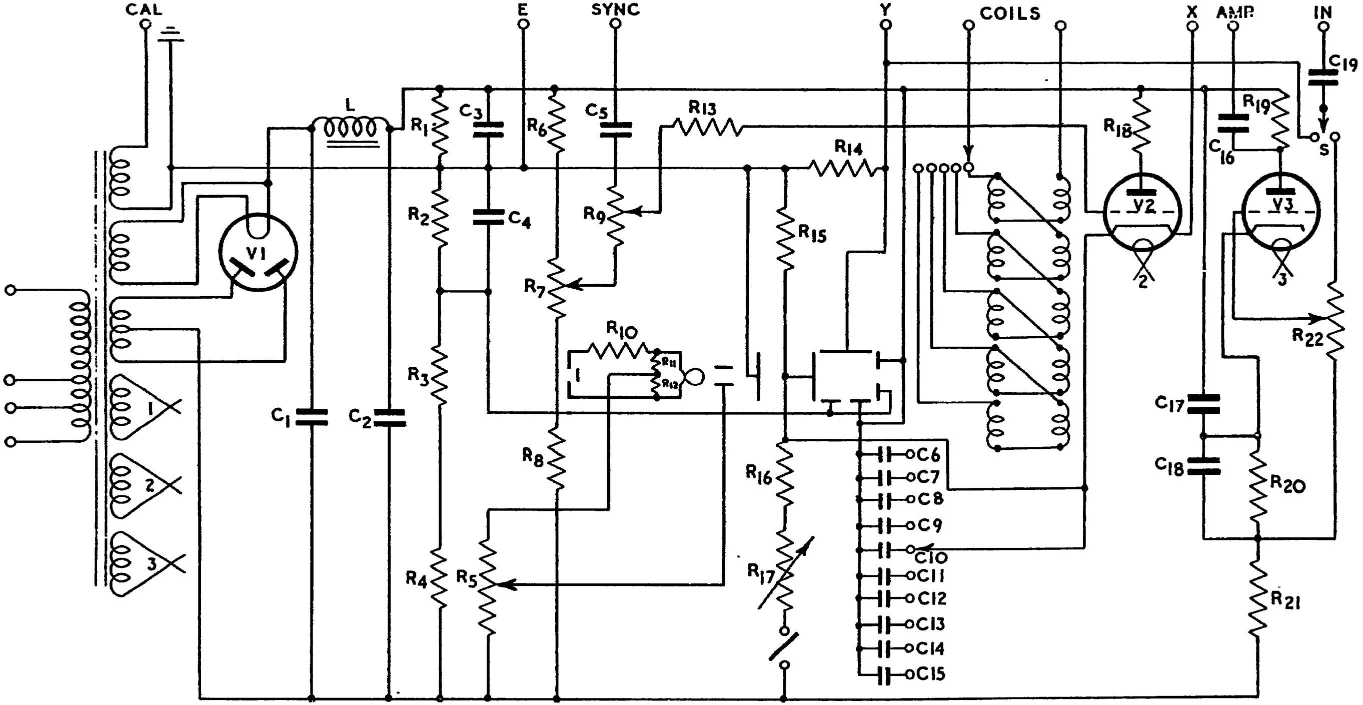

CAL
E
SYNC
Y
COILS
X
AMP
IN
C19
S
R19
C16
R18
V2
V3
2
3
R22
R13
R14
L
R1
C3
R6
C5
R2
C4
R9
R15
V1
R7
R10
R11
R12
1
C1
C2
R3
R8
R16
R4
R5
R17
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C17
C18
R20
R21
1
2
3

FIG. 17–33

may be found by counting the complete number of waves round the circular trace.

Oscillograph Construction

Where a cathode-ray tube is employed as an oscillograph the complete structure may be considered as being formed of three units: the tube, the time base, and the power supply. As most main supplies are a.c., the power unit must comprise a rectifier and suitable smoothing equipment. A transformer is necessary, possessing a number of secondary windings, several of which will supply the indirectly-heated cathodes of the valves and tube, while at least one winding will be employed in conjunction with a rectifier. From the latter is derived the tube anode supply, the time-base supply, various biasing potentials, and possibly the h.t. supply for a signal-amplifying circuit. The values of the various voltages are generally derived from resistance potentiometer networks, as indicated by Fig. 17–33. This circuit is that of an oscillograph designed and constructed by the author and may serve to illustrate design and construction procedure. It comprises a 4½ in. gas-filled tube, a gas-discharge time base, and a single-stage amplifier. The h.t. supply is at 850 volts and is smoothed by C_1C_2 and L. The tube is of split-plate construction to avoid origin distortion, the biasing potentials of $\pm$ 90 volts being derived from R_1 and R_2. The shield bias is of the order of $-$ 60 volts and is derived from R_5.

TIME BASE

The time-base condensers are charged through a variable resistance in series with a fixed resistance. The latter is 0·25 megohm, and the maximum value of the former is 2 megohms, so that a time-base frequency variation of 9 : 1 may be obtained with each condenser. From (17–16) we have

$$t = CR \log_e \frac{E}{E - V}$$

i.e. the time for the condenser voltage to rise from zero to V (where V is the thyratron discharge voltage) is proportional to the product CR. From this it appears that any combination of C and R might be employed for a desired value of t and time-base frequency. It must be remembered, however, that the charge on C is CV coulombs, and that this must be discharged through the thyratron once per cycle. If t_d is the discharge time, then the average current through the thyratron is CV/t_d and, in view of the small value of t_d, this may be detrimental to the life of the cathode. In view of this the largest condenser in the time-base group seldom exceeds 0·5 μF.

The bias for V_2 is derived from the potentiometer $R_6R_7R_8$ via the synchronizing potentiometer R_9. R_7 is the bias potentiometer and variation of this varies the length of scan that may be obtained. The initial value of the bias is approximately equal to the maximum value of the scan voltage, as there is only a difference of a few volts between the grid and cathode potentials of V_2 when a discharge occurs. The function of C_{19} is to absorb any d.c. component in the waveform under examination, and its impedance should be low compared with that of R_{14}. In order to effect synchronization, the terminal marked SYNC is connected to either *Y* or *IN*.

SIGNAL AMPLIFICATION

In many instances the amplitude of the waveform to be examined is too small for convenient observation. The size of the trace formed on the recording surface depends on the sensitivity of the cathode-ray tube, which may vary from 0·1 mm./volt with high-vacuum tubes to 2 mm./volt with those of the gas-filled type. Thus small voltages may need amplification, particularly when a high-vacuum tube is employed. Where the impedance of the work voltage supply is low the voltage may be increased by means of a transformer with a high primary impedance. In many instances, however, the supply has a relatively high impedance, and to avoid signal distortion such cases, of course, necessitate a valve amplifier. The amplifier must be free from amplitude and frequency distortion over a wide range and, to ensure this, where more than a single stage is employed, resistance/capacity coupling must be used.

Where a high degree of amplification is necessary, it is customary to employ a separate amplifier. In those cases where a moderate amount of amplification is sufficient it is possible and convenient to incorporate the amplifier with the oscillograph. Such is the case in Fig. 17-33, where a single-stage amplifier with a maximum gain of 20 is employed. It will be noted that the amplifier h.t. supply is derived from that of the tube and time base, which means that contrary to usual practice, the cathode of the amplifying valve is at a high potential to earth. A list of components for Fig. 17-33 is given by Table 17-1 on page 586.

Fig. 17-34 shows the circuit of a high-vacuum oscillograph designed by the author. This comprises a 3 in. dia. tube, a pentode-charged time base, and a single-stage amplifier. The tube is of the

TABLE 17–1

R_1	25,000 ohms	½ watt	C_1	2 μF	1000 VW
R_2	25,000 ohms	½ watt	C_2	2 μF	1000 VW
R_3	150,000 ohms	2 watts	C_3	8 μF	500 VW
R_4	30,000 ohms	½ watt	C_4	0·25 μF	500 VW
R_5	0·5 megohm	1 watt	C_5	0·02 μF	350 VW
R_6	30,000 ohms	½ watt	C_6	0·0001 μF	350 VW
R_7	50,000 ohms	1 watt	C_7	0·0003 μF	350 VW
R_8	130,000 ohms	1 watt	C_8	0·001 μF	350 VW
R_9	2 megohms	1 watt	C_9	0·002 μF	350 VW
R_{10}	1·2 ohms	2 watts	C_{10}	0·005 μF	350 VW
R_{11}	100 ohms	½ watt	C_{11}	0·01 μF	350 VW
R_{12}	100 ohms	½ watt	C_{12}	0·02 μF	350 VW
R_{13}	0·25 megohm	½ watt	C_{13}	0·05 μF	350 VW
R_{14}	5 megohms	½ watt	C_{14}	0·1 μF	350 VW
R_{15}	5 megohms	½ watt	C_{15}	0·25 μF	350 VW
R_{16}	0·25 megohm	1 watt	C_{16}	0·25 μF	500 VW
R_{17}	2 megohms	1 watt	C_{17}	4 μF	1000 VW
R_{18}	50 ohms	½ watt	C_{18}	8 μF	275 VW
R_{19}	100,000 ohms	1 watt	C_{19}	0·25 μF	1000 VW
R_{20}	1,000 ohms	½ watt			
R_{21}	40,000 ohms	½ watt			
R_{22}	2 megohms	1 watt			

VALVES

V_1 = Cossor 43IU
V_2 = Cossor GDT4B
V_3 = Mullard 354V

TUBE

Cossor 37J

TABLE 17–2

R_1	330,000 ohms	1 watt	C_1	2 μF	1000 VW
R_2	250,000 ohms	1 watt	C_2	2 μF	1000 VW
R_3	25,000 ohms	½ watt	C_3	0·1 μF	350 VW
R_4	33,000 ohms	½ watt	C_4	0·15 μF	350 VW
R_5	50,000 ohms	1 watt	C_5	0·03 μF	350 VW
R_6	2 megohms	1 watt	C_6	0·006 μF	350 VW
R_7	0·25 megohm	½ watt	C_7	0·0015 μF	350 VW
R_8	5 megohms	½ watt	C_8	0·0003 μF	350 VW
R_9	5 megohms	½ watt	C_9	0·00005 μF	350 VW
R_{10}	150,000 ohms	1 watt	C_{10}	2 μF	350 VW
R_{11}	250,000 ohms	1 watt	C_{11}	50 μF	25 VW
R_{12}	3,000 ohms	½ watt	C_{12}	0·25 μF	1000 VW
R_{13}	50 ohms	½ watt	C_{13}	0·5 μF	350 VW
R_{14}	100,000 ohms	1 watt	C_{14}	1·0 μF	500 VW
R_{15}	3,000 ohms	½ watt	C_{15}	1·0 μF	500 VW
R_{16}	2 megohms	1 watt			
R_{17}	133,000 ohms	2 watts			

VALVES

V_1 = Cossor 43IU
V_2 = Cossor GTD4B
V_2 = Mullard SP4
V_4 = Mullard 354V

TUBE

Mullard A40–G3/N3

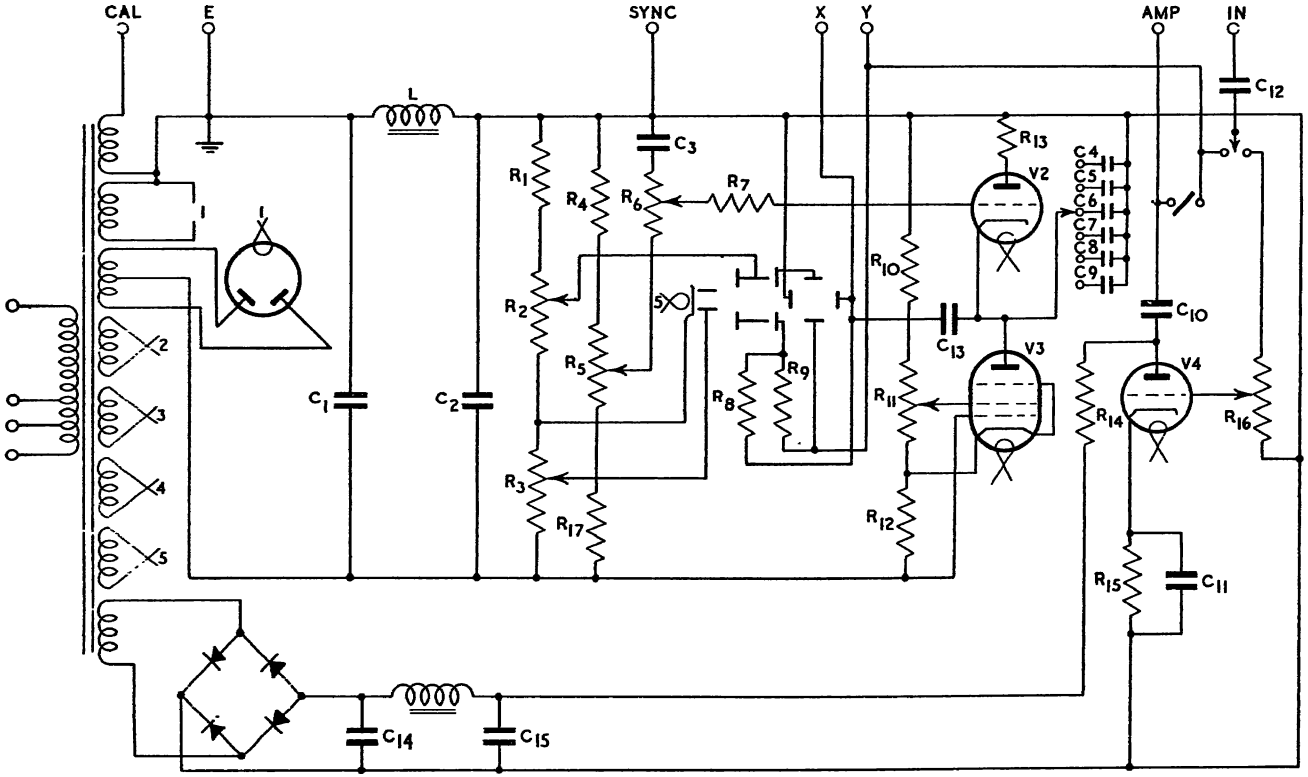

CAL
E
SYNC
X
Y
AMP
IN
L
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
R15
R16
R17
V2
V3
V4
1
1
2
3
4
5
5

FIG. 17–34

two-anode type with 250 volts on the first anode, 850 volts on the second anode, and 25 volts on the shield.

TIME BASE AND AMPLIFIER

The sensitivity of the tube is 0·15 mm. per volt for the Y-plates and 0·2 mm. per volt for the X-plates. Hence for a scan of 6 cm. a voltage variation of 300 volts is necessary. As this exceeds 20 per cent of the h.t. supply, it was necessary to employ constant-current charging through the pentode V_3. It will be noted that, in contrast to Fig. 17–33, the oscillograph of Fig. 17–34 possesses a separate h.t. supply for the amplifier, this supply being obtained by means of a metal rectifier. Fig. 17–35 shows the complete oscillograph, while Table 17–2 gives a list of the components employed.

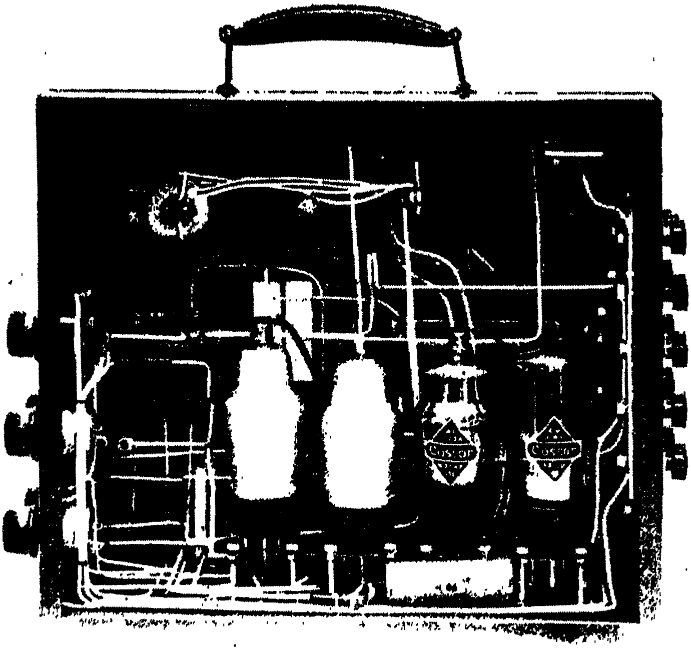


FIG. 17–35

CHAPTER XVIII

PHOTO-ELECTRIC CELLS

PHOTO-ELECTRIC cells fall under three headings: photo-emissive, photo-voltaic, and photo-conductive. Cells appertaining to the first heading function by virtue of electrons emitted from a metallic surface under the influence of incident radiation, the principles of this phenomenon having already been discussed in Chapter IV. Photo-voltaic cells operate by reason of an e.m.f. generated by physical or chemical reactions produced by the incident radiation, while photo-conductive cells undergo changes in electrical conductivity with variation in the intensity of the radiation. The three types of cells will be treated in the above-mentioned order.

Photo-emissive Cells

As these cells operate in accordance with principles laid down in Chapter IV, it is perhaps advisable before proceeding to recapitulate the salient points and to add to these where necessary—

1. The electrons are emitted from the cathode with a variety of velocities and may be completely extinguished by applying a retarding potential between the cathode and collecting anode. The potential must be such that the electric field is directed from cathode to anode. This means that the potential of the anode must be negative with respect to the cathode. The velocity and, of course, energy of the emitted electrons are independent of the intensity of the incident radiation, but are dependent on its frequency. The maximum energy with which an electron escapes is directly proportional to the frequency, as is also the value of the retarding potential to prevent escape.

2. When the anode is made a few volts positive, the photo-electric current reaches a saturation value. For radiation of a given frequency this value is proportional to the intensity of the incident radiation. This property is to be expected, for the number of electrons ejected per second will be proportional to the number of light-quanta, or photons, received per second.

3. The time lag between exposure of a photo-electric surface to radiation and the emission of electrons is negligible.

4. Every photo-electric material has a frequency and wave-length below and above which emission will not occur. These are known as the *threshold frequency* and *threshold wavelength*, and are

directly related to the work function of the material in the following manner.

$$h/\lambda_0 = hf_0 = E_w$$

where h is Planck's constant, f_0 and λ_0 are respectively the threshold frequency and wavelength, and E_w the work function.

5. The quantum efficiency or quantum yield of a photo-electric surface is the number of electrons emitted per photon. It is generally not more than a few per cent. The photo-electric efficiency is the amperes per watt of incident radiation and is sometimes referred to as the spectral sensitivity.

6. As the radiation frequency is increased above the threshold value the spectral sensitivity at first increases, then passes through a maximum and afterwards decreases. Typical curves for the alkali metals are shown by Fig. 18–1. The character of these curves may be explained in the following manner. As f increases above f_0, the energy per photon increases and with it the probability of electron emission. Thus, initially, the spectral sensitivity increases. However, if the total radiant power is W watts, the number of photons arriving per second is W/hf, this number decreasing with increasing frequency. Hence although the quantum yield may increase with frequency, it is limited as stated in 5 above. On the other hand, W/hf continually decreases with frequency and so leads to a maximum in the spectral sensitivity curve as shown.

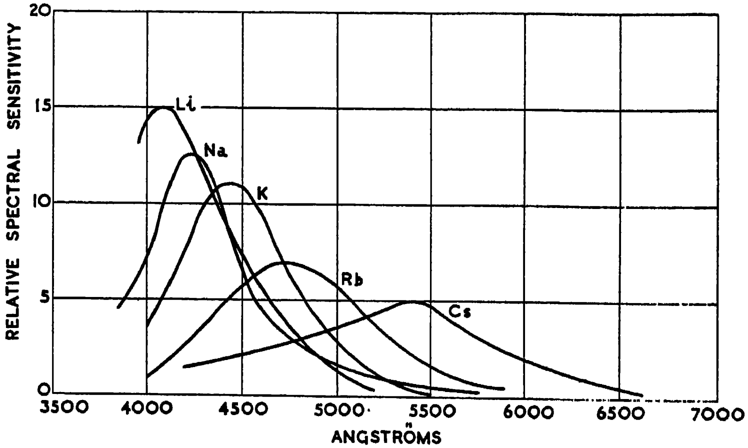


FIG. 18–1

7. For the majority of applications photo-electric emission may

be regarded as independent of temperature conditions. Actually, however, temperature has some slight effect, principally in modifying the threshold frequency. As given in 4 above, we have

$$f_0 = E_w/h$$

and evidently f_0 is only definite if E_w is constant. Now $E_w = E_0 - E_m$ and hence a sharp cut-off in the emission can only occur at 0° K. At room temperatures a few electrons exist within a metal with energies greater than E_m, these electrons obeying Maxwell's Law rather than that of Fermi-Dirac. Hence the energy required to release such electrons is less than E_w and thus some emission will occur for frequencies less than f_0. This explains why the curves of Fig. 18–1 decrease gradually to zero instead of exhibiting a sharp cut-off at f_0. Naturally this effect is more pronounced the higher the temperature becomes.

Vacuum Photo-cells

Practical forms of photo-emissive cells consist of a cathode and collecting electrode (anode) within a highly-evacuated glass or quartz bulb. In order that relatively large currents may be obtained, the cathode area is as large as possible, while the anode is small in order not to impede the reception of light or radiation by the cathode. A commonly employed electrode arrangement is a semi-cylindrical cathode with a coaxial wire anode. A typical cell of this type is illustrated by Fig. 18–2. The cathode itself is not constructed of photo-sensitive material, but a thin layer of this is deposited upon the cathode surface.

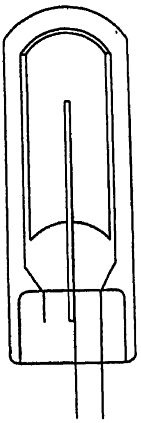


FIG. 18–2

VOLT-AMPERE CHARACTERISTICS

A set of typical volt/ampere characteristics for a vacuum type photo-cell is shown by Fig. 18–3. In order that no current shall flow it is, of course, necessary to apply a retarding potential, i.e. the anode must be negative with respect to the cathode. The characteristic is, of course, affected by the contact potential difference of the electrodes. For example, if the work functions of the cathode and anode surfaces are, respectively, 2·0 and 5·0 volts, then, in accordance with what has been stated on page 180, the anode surface is 3·0 volts negative with respect to that of the cathode when the supply potential difference is zero. The potentials shown by Fig. 18–3 are actual electrode surface potentials and do not refer to that of the supply. Now, theoretically, the emission current should saturate for zero potential difference but, as the characteristics indicate, saturation does not occur until the anode is some few volts positive. This is because of space-charge effects and the fact that the electrons are not necessarily emitted with purely radial velocities. Due to the latter, some electrons will strike the cathode surface and be re-absorbed. However, as the field strength and its directional influence are increased by raising the anode potential, practically the entire emission will be collected by the anode and saturation will result. It will be noted that even after saturation some slight increase in emission occurs, possibly due to a more complete collection of electrons and the influence of the Schottky effect.

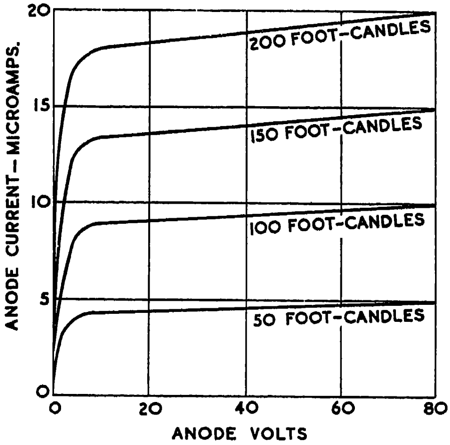


FIG. 18–3

Cathode Materials

It is evident from 4 above that the frequency range over which a photo-sensitive cell will operate depends on the work function of the cathode surface. As stated in Chapter IV, materials with a work function higher than 3 volts are useless for work within the visible range. Thus, as indicated by Table 4–1, most metals are unsuitable for photo-electric work as they only start to emit in the near ultra-violet part of the spectrum. If cells are to be used with invisible radiation, it is necessary to construct the bulbs, or windows through which the radiation passes, of quartz or special glass, as ordinary glass begins to absorb strongly at about 3300 Å. A spectral sensitivity curve for sodium is shown by Fig. 18–4, while relative

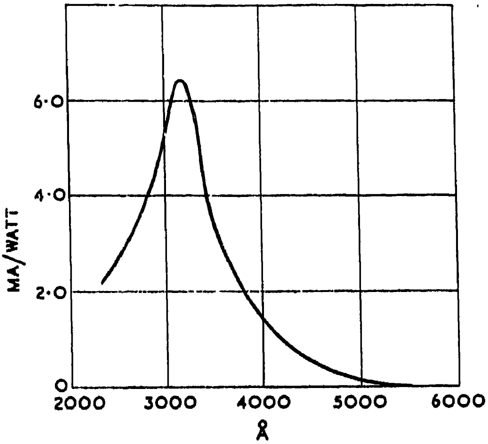


FIG. 18–4

values for various metals are given by Fig. 18–5. It may be noted that, generally speaking, the higher the value of λ_0, the higher is the maximum value of the sensitivity.

As stated in Chapter IV, it is essential that the metal surface should be uncontaminated if consistent results are to be obtained. However, consistency is greater with plain cathodes than with sensitized ones, the greatest consistency being obtained with the least electro-positive metals.

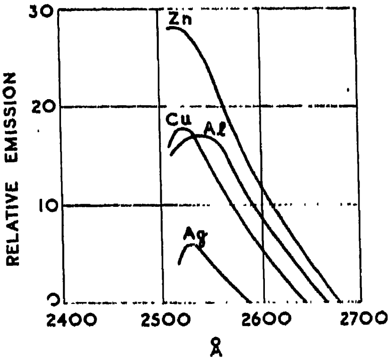


FIG. 18 5

SENSITIZED CATHODES

For operation in the visible range it is necessary to employ a cathode with a low-work function and this means an electro-positive metal such as potassium, rubidium, or caesium. However, these metals are particularly susceptible to a process known as sensitization, which shifts λ_0 further towards the red end of the spectrum and the maximum of the sensitivity curve towards the centre. Two processes of sensitization exist.

The earlier of the two consists of making the metal the cathode of a glow discharge in hydrogen. This has the effect of changing the appearance of the metal, giving potassium, for example, a matt surface and a blue-green colour. Fig. 18–6 shows spectral sensitivities curves for the alkali metals sensitized in the above manner. The second of the two methods of sensitization consists of exposing the metal for a short time to the action of vapours containing oxygen and/or sulphur. The effect of this is somewhat obscure but it is believed that compounds of the metals are formed and, subsequently, upon these is deposited a thin film of the uncombined metal. Fig. 18–7 shows sensitivity curves of metals sensitized by the latter process.

CATHODE FILMS

Instead of employing a cathode formed wholly from an alkali metal, it is possible to employ a relatively insensitive metal with a thin alkali film deposited thereon. With this arrangement it is found that as the thickness of the deposited metal increases from zero, the threshold wavelength shifts towards the red end of the spectrum until it is farther within the red than that of the deposited metal itself. Increasing the thickness of the deposit now commences to shift λ_0 towards the blue end until, when the deposit is very thick, the sensitivity curve is that of the deposited metal alone.

By depositing thin films of alkali metals on the oxides of relatively insensitive metals, cathodes may be produced which are more sensitive to red and white illumination than when the deposition is made on unoxidized metals. The process is effected by oxidizing a surface, say, silver, and then heating it to about 200° C. in a vapour of caesium, rubidium, or potassium. The result is generally represented by $x - o - y$, where x is the alkali metal, o the oxide, and y the insensitive metal. Thus a commonly employed cathode is given by $Cs - O - Ag$, which represents a very thin deposit of caesium on silver oxide produced in the above manner. Figs. 18–11, 18–12, and 18–13 refer to various cathodes of this type.

Gas-filled Photo-cells

The total emission current realizable in a vacuum cell is of a very small order, usually measured in microamperes. By admitting a small quantity of inert gas into the cell it is possible to amplify this current by means of ionization of the gas molecules. As shown in Chapter II the current through a gas when ionization by electron collision only occurs is

$$i = i_0 \varepsilon^{\alpha d} \quad . \quad . \quad . \quad . \quad (18–1)$$

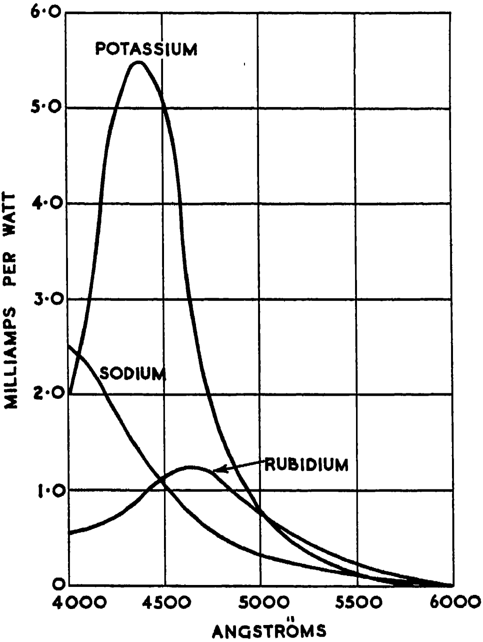


FIG. 18–6

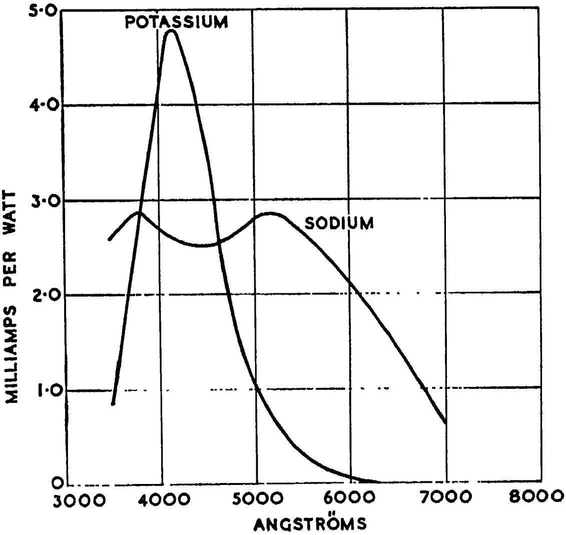


FIG. 18–7

In this case i_0 is the photo-electric emission current and d the distance between the electrodes. Hence the current in a gas-filled photo-cell is proportional to the photo-electric emission current i_0, and as this quantity is proportional to the radiation intensity, so also must be i. As previously shown (page 67),

$$\alpha = pf\left(\frac{X}{p}\right)$$

and thus i also depends on the gas density and field strength, the latter being proportional to the potential difference between the electrodes.

It is to be anticipated that below the ionization potential of the gas the characteristics of a gas-filled photo-cell are similar to those of a vacuum tube. This is so as is shown by Fig. 18–8, which gives characteristics of a gas-filled cell for various anode potentials and radiation intensities. Up to the ionization potential the curves obviously bear a close resemblance to those of Fig. 18–3. Above this (18–1) commences to operate, the current rapidly increasing with voltage due to the term α. It will be noted that i is approximately proportional to the radiation intensity. Of course, if the anode potential is sufficiently increased, a glow discharge will occur in accordance with the principles laid down in Chapter II for such discharges. This condition clearly must be avoided, for not only would such a discharge damage the photo-cell, but, as already shown,

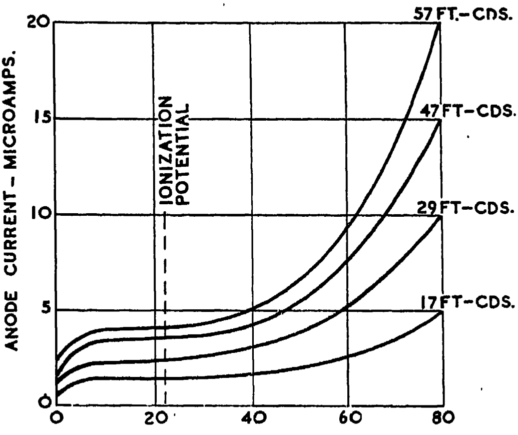


Fig. 18–8

once the discharge is started it is independent of the photo-electric current i_0 and thus also of the illumination intensity. In these circumstances, of course, the photo-cell no longer functions as such. Furthermore, as previously stated on page 80, when the photo-electric emission is sufficiently intense, it has the effect of lowering the breakdown voltage of a gas. Thus, it appears that the higher the illumination intensity the lower the breakdown potential of the cell. This is so and to every illumination intensity is an anode potential at which breakdown will occur. This is illustrated by Fig. 18–9, from which it will be noted that, for the photo-cell concerned, breakdown always occurs beyond about 120 volts, no matter what the value of the illumination intensity. Initiation of a glow discharge occurs for about 75 volts, and below this value a dark discharge only occurs irrespective of the value of the illumination intensity. Because of the possibility of a glow discharge the factor $\varepsilon^{\alpha l}$ in (18–1) is not allowed to exceed about 10 in practical cells.

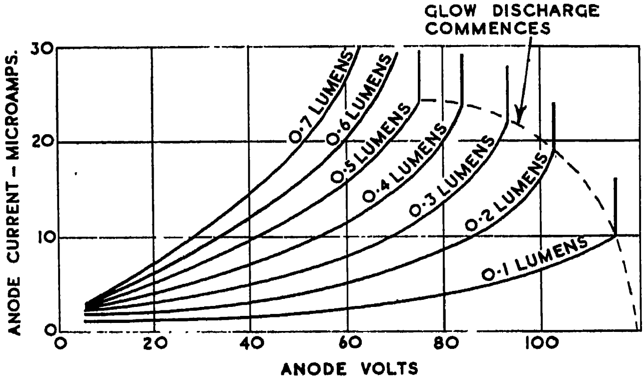


FIG. 18–9

SENSITIVITY

A definition of sensitivity has already been given under 5 above, this form being termed the *static* sensitivity of a cell. In a similar manner the *dynamic* sensitivity is the ratio of the alternating amperes per alternating watt of incident radiation. In the case of vacuum photo-cells these two sensitivities are equal. In the gas-filled cell, however, the dynamic sensitivity is a function of the frequency, decreasing with an increase in the latter. This is due to the finite amount of time necessary for the development of the ionization current in the cell. A typical instance of the variation of dynamic sensitivity with frequency for a gas-filled cell is shown by Fig. 18–10. It will be noted that such cells are, generally speaking, only suitable for use within the audio-frequency range. Actually they find considerable employment in connexion with talking films.

The equation of (18–1) shows that i is proportional to i_0. This, however, is only true while α is constant. According to (18–1), for a constant value of p, α is a function of X and thus of the electrode potential difference. In practical photo-cell circuits the cell is usually in series with an impedance, with the result that a change in illumination intensity changes the voltage across the cell. Thus, if the illumination intensity is increased, the photo-electric current is also increased but, at the same time, the electrode potential difference is decreased. This causes a decrease in α and thus a decrease in the amplifying factor $\varepsilon^{\alpha d}$. Hence two opposing factors are at work and the result is a non-linear current/illumination characteristic. In the case of a vacuum cell this does not occur, providing the electrode potential difference does not fall below that necessary to produce the saturation value of i_0. It will be appreciated from the foregoing that gas-filled cells may suffer both from frequency and amplitude distortion, and hence where almost

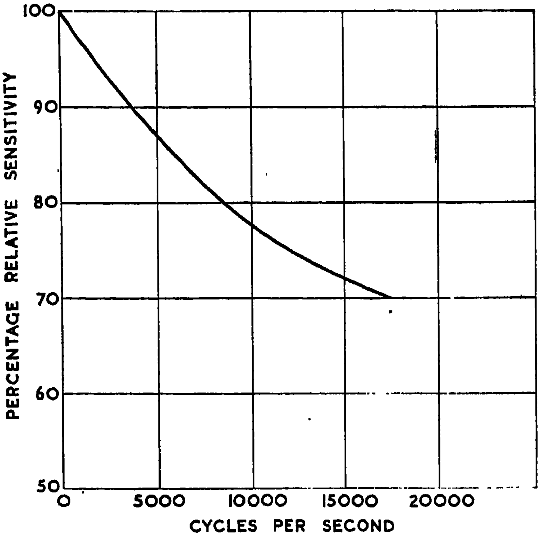


Fig. 18–10

complete freedom from these defects is desired it is necessary to employ a vacuum cell.

SPECTRAL RESPONSE CURVES

In practical applications it is frequently desirable to employ a cell whose frequency response curve is similar in form to the energy-distribution curve of the source of radiation. As indicated by Figs. 18–1, 18–6 and 18–7, various response curves may be obtained by employing various materials for the cathode surface and subjecting these to different processes.

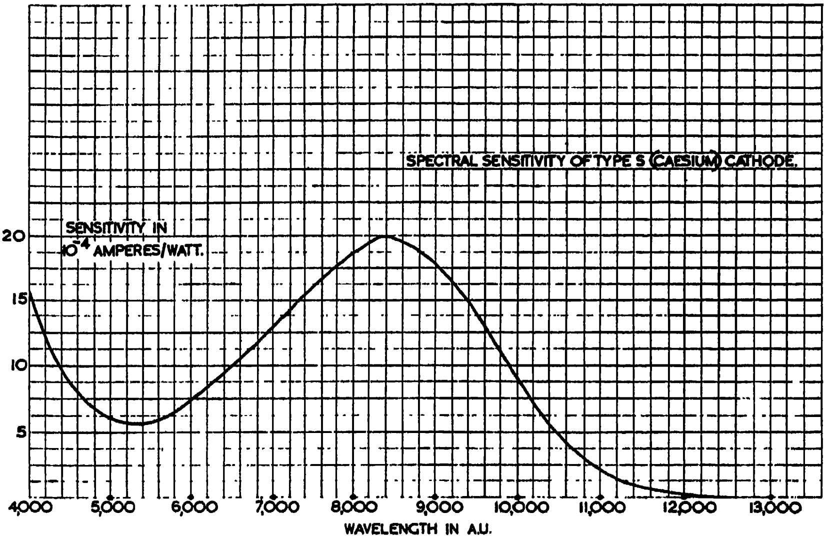


FIG. 18–11

By courtesy of Cinema-Television, Ltd.

Some typical response curves of cells manufactured by Cinema-Television, Ltd., are shown by Figs. 18–11, 18–12, and 18–13. It will be noted the type *S* cell has a peak sensitivity in the near infra-red, while the maxima of the types *A* and *B* are near the blue end of the spectrum. Furthermore, it is apparent that the sensitivity of the *A* cell at its maximum is some thirty times greater than that of the *S* cell. To render the difference in sensitivity more evident, the curve of Fig. 18–11 is shown again in Fig. 18–12. The outstanding superiority of the *A* cell compared with the *S* cell is not

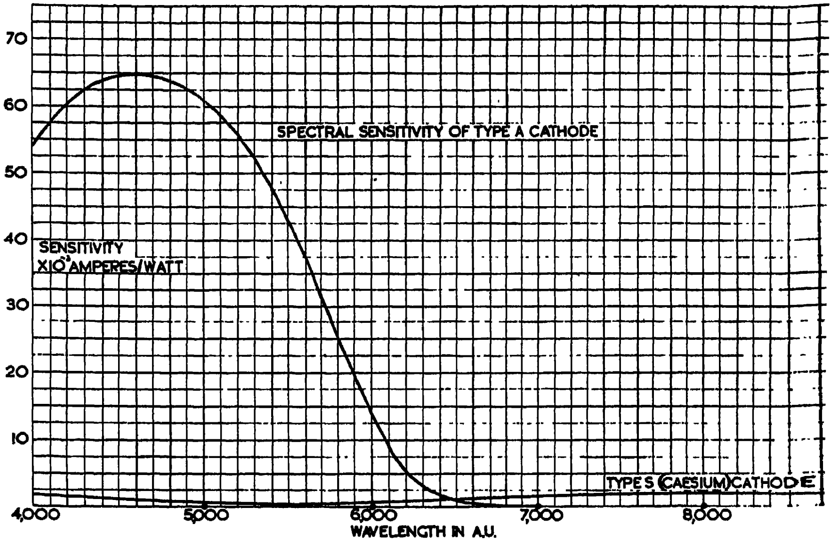


FIG. 18–12
By courtesy of Cinema-Television, Ltd.

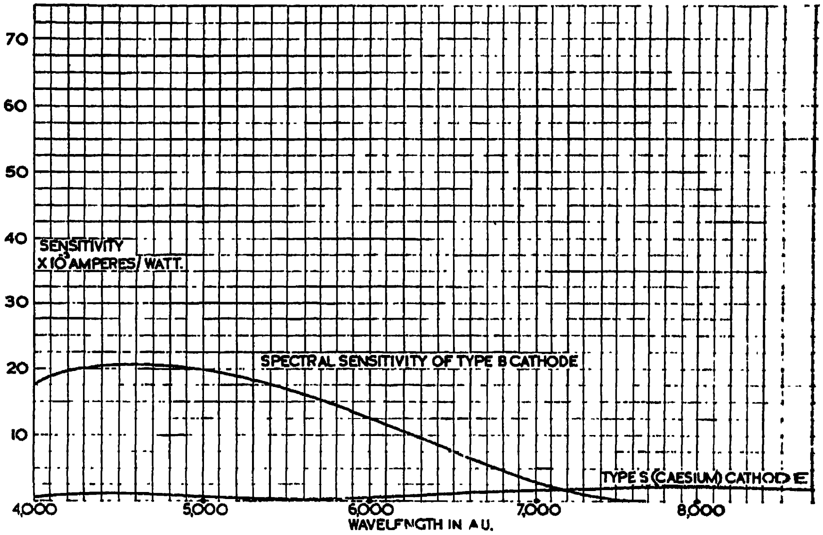


FIG. 18–13
By courtesy of Cinema-Television, Ltd.

so marked if the light source consists of an incandescent lamp, because the latter has a maximum light emission in the infra-red and a very small emission in the blue part of the spectrum. However, even with the light of an incandescent lamp the type *A* cell still shows an improvement over the type *S* cell. The type *B* cell has a peak sensitivity in the same range of the spectrum as the type *A* cell, but while the maximum value is lower, the sensitivity extends further into the red part of the spectrum.

In selecting either of the three cells of Figs. 18–11, 18–12, and 18–13 for a particular application, the following points should be considered—

1. If infra-red light is employed, obviously only type *S* is suitable.
2. With an incandescent light source type *A*, despite the unfavourable position of its peak sensitivity, is about twice as sensitive as type *S*.
3. If sensitivity to daylight is required, type *A* is from ten to twenty times superior to type *S*, because the peak emission of the sun is much nearer the blue end of the spectrum than that of an incandescent lamp.
4. If a photo-cell is required with a sensitivity similar to that of the human eye, type *B* is most suitable.
5. In some instances maximum sensitivity to the light of a cathode-ray tube fluorescent screen is required. In these circumstances the choice of photo-cell naturally depends on the colour of particular fluorescent screen employed. Where a screen with a blue-green fluorescence is used, the type *A* cell is more sensitive than the other two types. Where negligible afterglow is essential, screens with this property have an almost pure blue emission, in which case the response of an *A* type cell may be as much as 200 times greater than that of an *S* type cell.
6. For ultra-violet radiation the type *A* cell is suitable, being sensitive down to 2000 Å.

Photo-cell Applications

In order to indicate the effects of radiation upon a photo-cell the basic circuit of Fig. 18–14 is generally employed, where *R* is a non-inductive resistance. In the case of a vacuum cell, providing the cell voltage is always above the value necessary to produce saturation, the current in

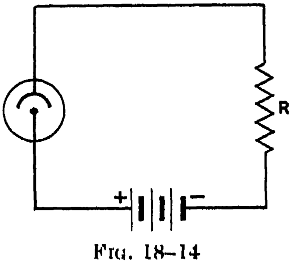


FIG. 18–14

R is proportional to the illumination intensity. The voltage on R is $v = i_0R$ and hence v also varies directly with the illumination intensity. Thus it will be appreciated that the output voltage across R is free from amplitude distortion.

In the event of the volt/ampere characteristics being non-linear, the voltage across R is no longer a linear function of the illumination intensity. In these circumstances the value of i_0 and v must be derived in the following manner. The voltage across the cell for given values of E, i_0, and R is

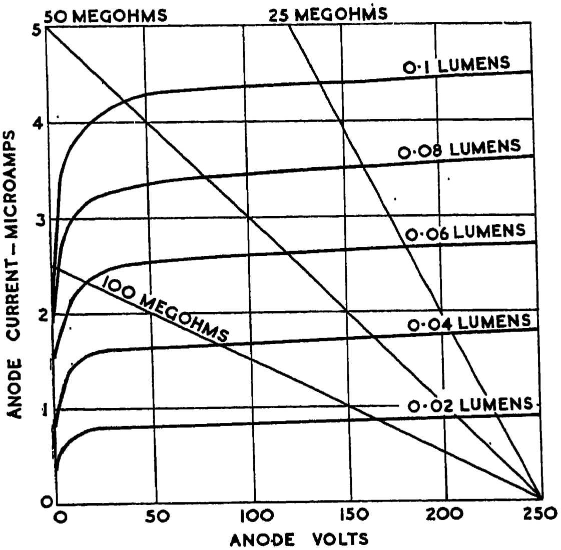


FIG. 18–15

$$v_a = E - i_0R$$

and

$$i_0 = \frac{1}{R}(E - v_a)$$

the latter being the equation to a straight line the tangent of which is $-1/R$. This line makes intercepts on the v_a and i_0 axes of E and E/R, respectively. The line, i.e. the load line, is now drawn across the characteristics as shown by Fig. 18–15 and from this construction i_0 and v_a may be derived for any value of R. Three values of R are given and the derived values of i_0 are shown by Fig. 18–16. It will be noted that while the current/illumination intensity relationship is approximately linear for resistance values of 25 and 50, a serious departure from linearity occurs for 100 megohms. It is evident that if non-linearity is to be avoided the value of the resistance must be such that the load line does not intersect the sensibly non-linear portions of the characteristics of Fig. 18–15. This, of course, tends to set an upper limit to the value of R. On the other hand, the output voltage is proportional to R and from this viewpoint it is desirable that R should be as large

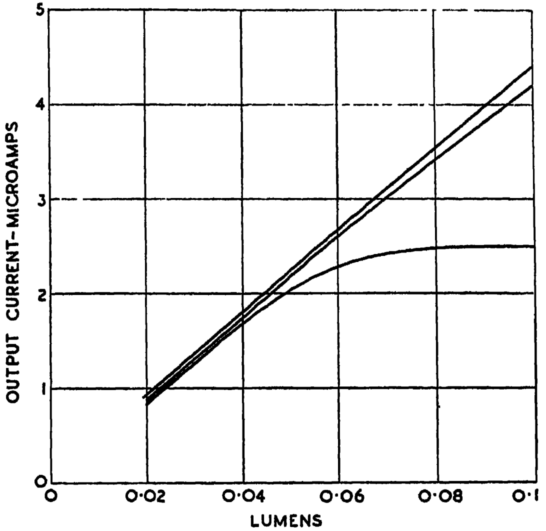


FIG. 18–16

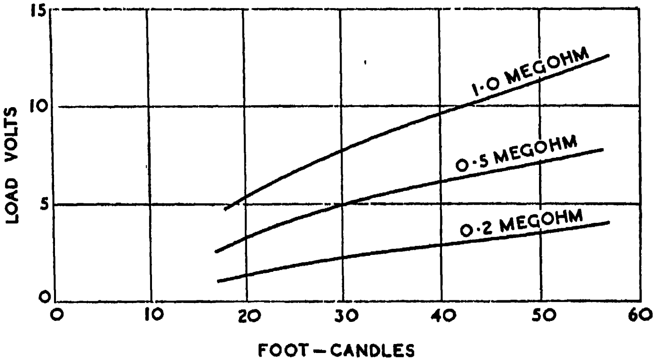


FIG. 18–17

as possible. However, apart from considerations of linear response, an upper limit is set for R by the leakage resistance of the photo-cell and the input impedance of the customarily employed amplifier.

The characteristics of Figs. 18–15 and 18–16 concern a vacuum cell. In the case of a gas-filled cell it is, of course, always necessary to draw the load line to derive the output characteristic, because the current/illumination intensity characteristics are continuously curved. Reference to Fig. 18–8 shows that the vertical intervals, for any given constant anode voltage, are approximately equal. Hence for $R = 0$ the current/illumination characteristic is linear. For $R \neq 0$ the vertical intervals decrease in magnitude as R increases, resulting in a non-linear output characteristic. Hence in order that the output characteristic shall not depart too seriously from linearity, it is desirable that R shall be as low as possible consistent with other requirements, such as high amplification. Fig. 18–17 shows output characteristics derived for three different values of R. It will be noted that as R increases so also does the departure from linearity.

The Multiplier Photo-cell

As previously stated, the total emission current realizable in a vacuum cell is of a very small order, and in some cases currents as low as 10^{-8} amp. are involved. One method of amplifying this current is, as we have already seen, by means of gas amplification. However, the amplification due to this method cannot, generally, be greater than ten because of the possibility of a glow discharge developing and the relatively poor frequency response of a gas-filled cell. With both the vacuum and gas-filled cells it is invariably necessary to amplify the output still further, this involving the employment of a multi-stage valve amplifier connected to the resistance R of Fig. 18–14.

In order to overcome the limitations of the simple vacuum and gas-filled cells, a multiplier photo-cell may be employed. In this cell amplification is achieved by employing the phenomenon of secondary emission, described in Chapter IV. The photo-electrons emitted from the cathode do not proceed directly towards the anode, as with a simple cell, but are directed towards a secondary emitting electrode maintained at a positive potential lower than that of the anode. At this electrode secondary electrons are generated which are greater in number than the primary photo-electrons responsible for their genesis. These secondaries may then be directed to a further secondary emitting surface, where they will again be multiplied. It is apparent that this process may be repeated many times by employing a number of electrodes each of which is at a higher potential than that from which it receives electrons. The total amplification obtained depends on the number of secondary emitting electrodes employed and on the secondary emission ratio of each. With regard to the latter quantity, as we have seen in Chapter IV, its value depends on the nature of the emitting surface and the energy of the primary electrons. For a suitable photo-sensitive surface and a primary electron energy of about 500 volts, the secondary emission ratio may be as high as 8; i.e. each primary electron releases eight secondary electrons. If the ratio is denoted by S and there are n stages of multiplication, it is clear that the overall electron amplification is S^n. In practice the value of S^n may easily be as high as 10^6.

In order to illustrate the process of electron multiplication the elementary diagram of Fig. 18–18 may be considered. Here light incident upon the cathode releases electrons photo-electrically. These electrons are directed by an electric field towards the electrode A, where they release further electrons by secondary emission. The latter electrons are now similarly directed to B, where a similar process occurs. This process is then continued until the anode is reached. If i_0 is the initial photo-electric current, then the final current is given by

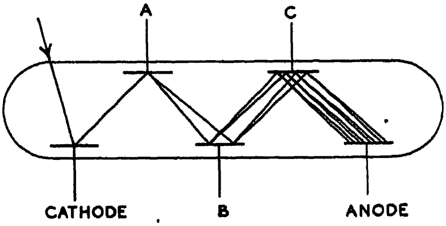


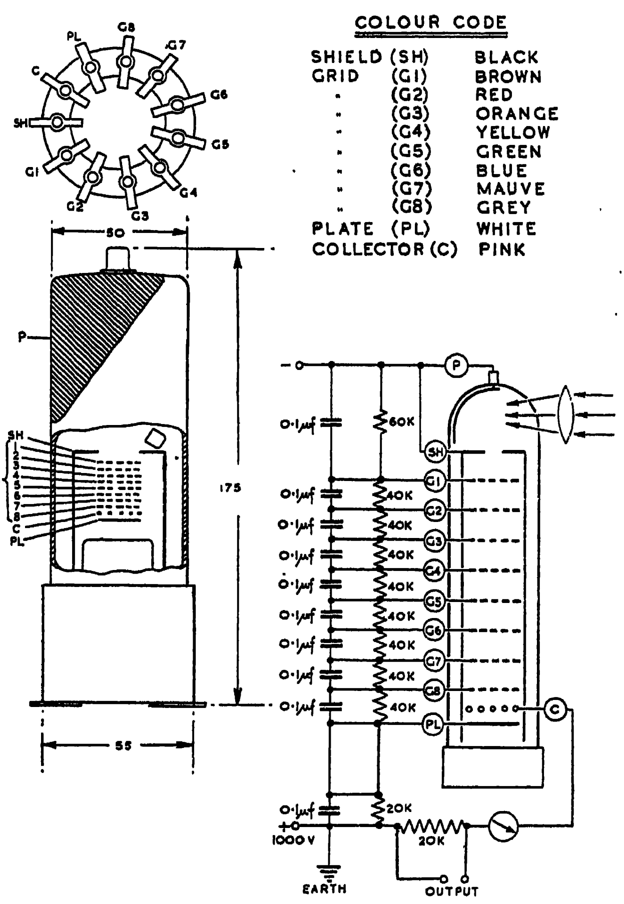
Fig. 18–18

$$i = i_0 S^n$$

The simple arrangement of Fig. 18–18 merely illustrates the principle of electron multiplication. In practice, difficulties exist in producing a multiplier with many stages of multiplication because, unless careful consideration is given to the geometry of the electrodes determining the electron paths, the electrons tend to by-pass the intermediate stages and travel more or less direct to the anode with reduced multiplication or even an absence of this. These difficulties have been successfully met by Cinema-Television, Ltd., with their "Baird" multiplier photo-cell, which is illustrated by Fig. 18–19. For the following description of the cell the author is indebted to its manufacturers mentioned above.

P is the photo-electric cathode which is deposited on the glass wall of the cell. The multiplier system is enclosed in an electrostatic shield *Sh* of cylindrical shape, which should be connected externally to *P*, i.e. to zero potential. Behind a circular aperture in the shield

MULTIPLIER PHOTO CELL FIG.B.



COLOUR CODE

SHIELD	(SH)	BLACK
GRID	(G1)	BROWN
"	(G2)	RED
"	(G3)	ORANGE
"	(G4)	YELLOW
"	(G5)	GREEN
"	(G6)	BLUE
"	(G7)	MAUVE
"	(G8)	GREY
PLATE	(PL)	WHITE
COLLECTOR	(C)	PINK

RESISTANCE POTENTIOMETER FOR SUPPLYING ELECTRODE POTENTIALS TO MULTIPLIER PHOTOCELL FROM 1000 VOLT RECTIFIER SUPPLY

FIG. 18–19

By courtesy of Cinema-Television, Ltd.

is a series of eight fine mesh grids G_1 to G_8, which each have a surface of high secondary emission factor S. G_1 is at a positive potential with respect to P, G_2 is more positive than G_1, G_3 is more positive than G_2, and so forth.

The action of the multiplier is as follows—

The photo-electrons released from P under the incidence of light enter the multiplier and will either impinge on the solid part of the grid G_1 with subsequent release of secondary electrons, or pass through the interstices. The secondary electrons released from G_1 will be pulled through the interstices towards the grid G_2, where the process is repeated, i.e. some (about half) of the electrons will be multiplied by secondary emission while the remainder will pass through the interstices unchanged. The result is that if the surface of the grid has a multiplication factor S of, say, five times, the apparent factor is only about 2·5. The same process takes place at the subsequent grids G_3 to G_8. The electrons leaving the last grid G_8 are made to pass through the wide mesh grid C to the secondary electron emitting plate Pl, which is more positive than G_8. Pl is the only electrode from which the full factor S is obtained. Finally, the electrons leaving Pl will be collected by the collecting electrode C, which is positive with respect to Pl and, of course, the most positive electrode of the whole system.

The total multiplication of the multiplier varies between 1000 and 5000 if an overall voltage of 1000 volts is applied with suitable subdivision for the individual grids. The recommended potentials for each electrode are as follows—

Cathode	P	0
Shield	Sh	0
Grid	G_1	150
„	G_2	250
„	G_3	350
„	G_4	450
„	G_5	550
„	G_6	650
„	G_7	750
„	G_8	850
Plate	Pl	950 ± 30
Collector	C	1000

As shown in the table, it is preferable to have the potential of Pl variable ± 30 volts so that it can be adjusted to the most suitable value; this is dependent on the maximum output current, i.e. the larger the current collected at C, the greater should be the potential difference between Pl and C. The voltages applied are not critical because the factor S changes only slightly with potential, for instance a change of ± 5 volts at one grid with respect to a neighbouring grid has no noticeable effect. The overall voltage can be raised to

1500 volts (corresponding to 150 volts between successive stages) or reduced to 500 volts (corresponding to 50 volts per stage), but 1000 volts (100 volts per stage) seems to be the most favourable potential because S decreases much more rapidly when the voltage per stage is dropped from 100 to 50 volts than it increases when it is raised from 100 to 150 volts.

In practice the overall voltage is usually supplied by a mains rectifier unit capable of providing an output of 5 mA at 1000 volts. The potentiometer providing the potentials for the individual stages should take a continuous load of 2·5 mA so that the current taken by the multiplier does not disturb the potential distribution. It is also advisable in many cases to insert a decoupling condenser between successive stages.

A satisfactory method of connexion is shown in the diagram of Fig. 18–19, in which the values given for resistances and condensers are those used for the detection of television signals. Obviously the most suitable values depend on the particular application and the diagram should only be regarded as one example.

Whether a multiplier cell or a simple cell with valve amplification is to be employed for a particular application is evidently a matter for some consideration. By employing a multiplier cell, valve amplification may still be necessary, but, of course, it is possible to cut this down by several stages. This results in an economy in amplifier cost and space, which, however, tends to be counterbalanced by the increased cost of the multiplier cell and the high-voltage supply needed for its operation. If the equipment with which the cell is to work already possesses a suitable supply, then the second of the two disadvantages mentioned is largely removed.

As the electrons proceed from electrode to electrode the heating effects at the electrodes clearly increase. Because of this it is usual to limit the current at the anode to 1 mA. Assuming an overall amplification of 10,000, this means that the original photo-electric current must not exceed 0·1 μA. If the cathode sensitivity is 30 μA per lumen, this means that the illumination must not exceed about 0·003 lumen. It is the definite inability of the multiplier photo-cell to handle only small illumination, that sets a limit to its applications, for if the illumination is likely to exceed the maximum permissible figure the amplification must be reduced. If this is done, then the cell may not offer any advantages over the simple cell with a valve amplifier. In the case of sound-film reproduction the maximum illumination likely to be handled is of the order of 0·1 lumen, which corresponds to a primary photo-electric current of 3 μA for a cell having the sensitivity given above. Hence in this case an electron multiplication of about 350 must not be exceeded if the anode current is not to be in excess of 1 mA. This degree of amplification is such as hardly to justify the use of a multiplier cell.

From the foregoing it appears that the field of application of the multiplier cell is the measurement of very small degrees of illumination. In such instances the employment of a simple cell necessitates the measurement of very small currents of the order of 10^{-9} amp. These currents may be directly measured or a valve amplifier employed. However, by means of a multiplier cell the difficulties of low current measurement may be avoided and highly sensitive galvanometers and valve amplifiers dispensed with.

FREQUENCY RESPONSE

The frequency response of a multiplier cell is of the same character as that of a simple cell. However, if the latter is employed with a valve amplifier, the amplifier is liable to introduce some frequency distortion. Hence from the high-frequency viewpoint a multiplier cell is to be preferred to the simple cell accompanied with valve amplification.

SIGNAL-TO-NOISE RATIO

The comparison of the multiplier cell with a simple cell with regard to unwanted components in the output is important. The unwanted or "noise" components arise from two causes: (1) the random electron emission from the cathode, and (2) from the amplifier. If the noise originates in the amplifier, then the employment of a multiplier cell instead of a simple cell and amplifier may be advantageous. However, if the noise comes under heading (1), it will be amplified by the multiplier cell, and the employment of this type of cell will not show any advantage. Where considerable interference is likely to originate from an amplifier, the employment of a multiplier cell is indicated, and in this case the smaller the noise originating from cause (1) the greater will be the improvement over the simple cell with amplifier.

As the photo-electric emission of a cathode decreases, the signal-to-noise ratio also decreases. This, of course, means that there is a relative increase in the random to photo-electric emission which will ultimately set a lower limit for the signal amplitude. This limitation is the reason why multiplication factors of more than 5000 are seldom employed.

In some photo-cell applications there is a constant background illumination upon which a varying component is superposed. This has the effect of decreasing the signal-to-noise ratio because the background increases the noise, but contributes nothing to the signal. It follows that it is desirable for the varying component to produce a high degree of modulation of the background illumination, for the lower the degree of modulation the lower is the signal-to-noise ratio.

In addition to the photo-electric emission of a cathode there exists a small amount of thermionic emission which makes its contribution to the noise problem. It follows that it is desirable to employ a cathode surface where thermionic emission is as low as possible, and this may be effected by the employment of an antimony-caesium cathode.

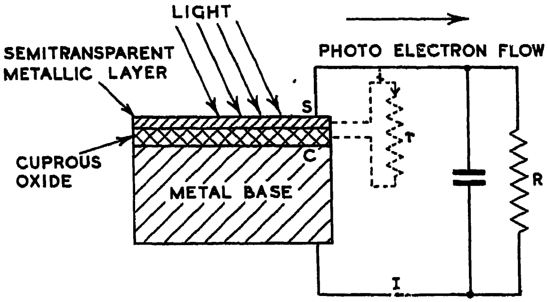


Fig. 18–20

Photo-voltaic Cells

Under suitable conditions the copper-oxide and selenium rectifiers described in Chapter XIII may be employed as photo-electric cells. Such cells are termed *photo-voltaic* and depend for their action on the generation of an e.m.f. within the cell by the incident radiation. Thus no battery or other source of supply is necessary for the operation of the cell and this, in certain applications, is a considerable advantage. The copper-oxide cell is similar to the copper-oxide rectifier and consists of a copper base upon which is formed a layer of cuprous oxide (Cu_2O). Upon this layer a semi-transparent metallic layer is formed (the counter-electrode) by cathodic sputtering. Contact is made with the counter-electrode by means of a thin copper ring, the complete arrangement being shown by Fig. 18–20. For obvious reasons this type of photo-electric cell is sometimes termed a "sandwich" or "rectifier-type" cell. The selenium cell consists of an iron base, upon which is deposited a thin layer of iron selenide, this being covered with a semi-transparent counter-electrode of silver. The frequency response of both cells is similar, the maximum response occurring approximately in the middle of the visible spectrum. However, the response of the selenium type is wider than that of the copper-oxide, the latter not extending beyond the visible range. A typical frequency response curve for a copper-oxide cell is shown by Fig. 18–21, and its resemblance to the relative sensitivity curve of the human eye is evident.*

Following the foregoing brief description of the photo-voltaic cell, we must now consider its structure in greater detail. Actually two types of cell exist, termed *back-wall* and *front-wall* cells, according to the manner in which the barrier or boundary layer is formed. In the first type the electrons within the cell travel in the same direction as the light, whereas in the second type they travel in the opposite direction. In the case of the copper-oxide cell, if a simple boundary is formed by pressing a thin metal film on to the oxide surface, the barrier layer or active boundary occurs at the junction of the copper-oxide and the parent copper. This cell is of the back-wall type and the oxide must be thin in order that light may penetrate to the barrier layer. If, however, a metallic film is sputtered on the exposed surface of the oxide, reduction of the latter will occur and a barrier layer will be formed thereat. In this case a front-wall cell is formed. As the barrier layer is always located between the semi-conductor and the sprayed counter-electrode in the selenium rectifier, it follows that selenium cells are always front-wall cells.

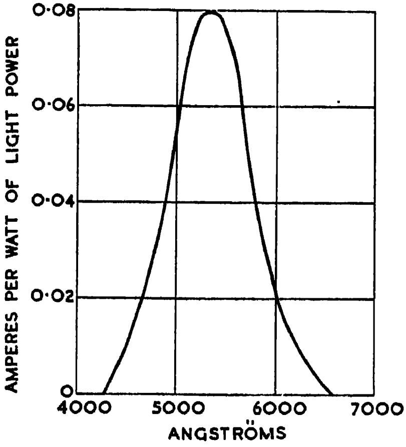


FIG. 18–21

Upon exposing the above-described cells to radiation of suitable frequency, an e.m.f. is generated due to the radiant energy causing electrons to pass from the semi-conductor (i.e. either the copper-oxide or the iron selenide) to the counter-electrode. This e.m.f. rapidly reaches a limiting value because the accumulation of electrons in the counter-electrode sets up a field opposing the electron motion towards this electrode. Furthermore, the electrons tend to flow back internally from the counter-electrode to the semiconductor. The relationship between the intensity of the incident radiation and the generated voltage may be expressed by the empirical formula

---

* See Fig. 19–1.

$$v = aJ^n$$

where J is the radiation intensity and a and n are constants, the latter being equal to about 0·4. It might appear that v should vary directly with J. However, in dealing with metal rectifiers it was found that the internal resistance decreased with an increase in voltage across the rectifier. Hence, as v is the product of the photo-electric current and the internal resistance, this voltage will be less than proportional to J.

Referring to Fig. 18–20, the external resistance and current of the photo-voltaic cell will be respectively denoted by R and I and the internal resistance and current by r and i. The latter is usually referred to as the primary current. Considering the cell as being of the front-wall type, the electrons flow from S to C, and thus the primary current i has the same direction as a "reverse" rectifier current. As with a rectifier, r is a decreasing function of i and hence r decreases with the illumination intensity. If R is not equal to zero, the volt drop across the cell is RI and this tends to produce a "forward" rectifier current through the cell equal to $(i - I)$. Hence

$$r(i - I) = IR$$

and

$$I = i\frac{r}{R + r} \quad . \quad . \quad . \quad (18\text{–}2)$$

Now i varies directly as the illumination, while r is a decreasing function of this quantity. Hence I is a non-linear function of the illumination as shown by the curves of Fig. 18–22. It is evident from (18–2) that if R is low compared with r, I is then approximately proportional to i and hence to the illumination. This is, of course, also evident from Fig. 18–22. R is usually the resistance of some indicating instrument (a milliammeter), and when this is very low the cell may be regarded as short-circuited. Under these circumstances the relation between the photo-electric current and the illumination intensity is linear as shown.

The lack of linearity in the curves of Fig. 18–22 is due to the presence of r and hence the value of this quantity should be as high as possible. As r varies inversely as the area of the cell, it is desirable to have the cell area no larger than that exposed to the incident radiation.

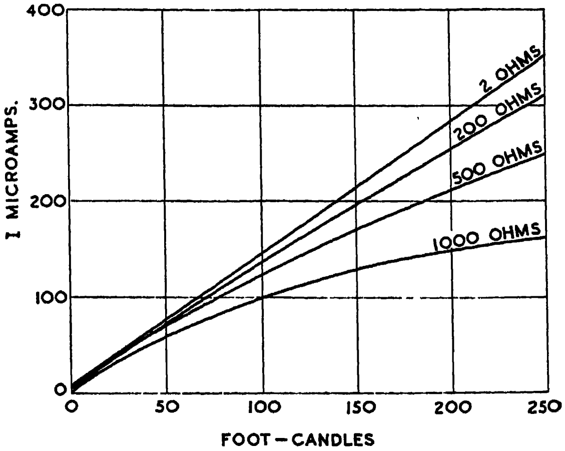


Fig. 18–22

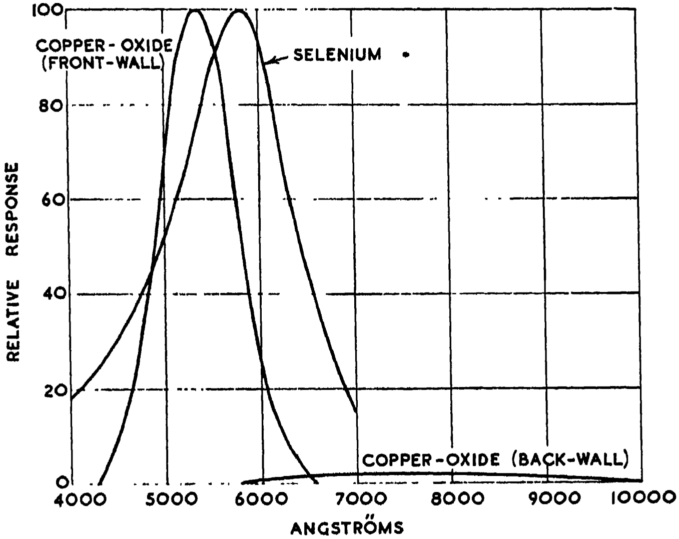


Fig. 18–23

SPECTRAL RESPONSE

Some spectral response curves for various photo-voltaic cells are shown by Fig. 18–23, these being obtained with a low value of R so that $I \fallingdotseq i$. The difference in the response of the front-wall and back-wall cells is immediately apparent and is due to the absorption of light by the cuprous oxide layer. For this reason back-wall cells are seldom employed.

Comparing Figs. 18–3 and 18–22, it is evident that the quantum and photo-electric efficiencies of photo-voltaic cells are much higher than those of photo-emissive cells. However, the former do not lend themselves to amplification by valve amplifiers as do the latter.

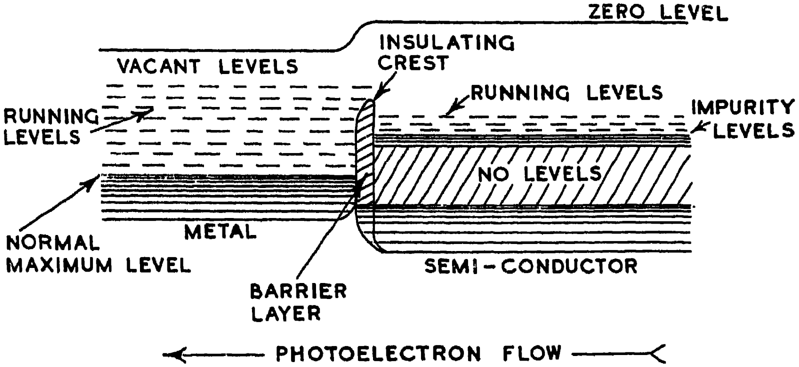


FIG. 18–24

THEORY OF CELL OPERATION

As explained in Chapter IV, a photo-electric effect occurs with insulators, but the threshold frequency is much higher than with metals. In the case of a semi-conductor the photo-electrons are presumably derived from the impurity levels (Chapter IV, page 196). Reference to Fig. 18–24 shows a theoretical construction of a photo-voltaic cell. It consists of a good conductor and semi-conductor separated by an insulating barrier layer, which is perhaps no thicker than 10^{-6} cm. Because of this layer, the normal maximum levels of the conductor and semi-conductor are not aligned as they are with two contacting conductors. Hence the potential barriers unite to form a crest at the insulating layer as shown. At the threshold frequency it may be presumed that the energy of the incident photons is such as to impart to the electrons in the impurity levels just sufficient energy to enable them to surmount the potential crest. Hence photo-electrons will pass from right to left, i.e. from the semi-conductor to the good conductor. Due to the lower normal maximum level in the good conductor electrons from this cannot pass to the semi-conductor. However, if the frequency of the incident radiation is continually increased towards and beyond the violet end of the spectrum, it is possible to cause electrons from the good conductor to pass over the potential crest to the semi-conductor. Ultimately a condition may be reached at which the net electron transfer over the crest is zero. This, of course, corresponds to zero response at the higher end of the spectrum in Fig. 18–21.

EFFECT OF TEMPERATURE

As with a rectifier, an increase in temperature decreases the value of r for the photo-voltaic cell. Referring to (18–2), if R is not small compared with r, then an increase in temperature will result in a decrease in I. If, however, R is negligible compared with r, temperature variations may have little or no effect on the cell response. These considerations, of course, presume that i remains constant. At very low temperatures it is probable that i decreases because of the relatively few electrons existing with energies in excess of the normal maximum level of the impurity levels.

Tests on various cells show that for the copper-oxide cell I decreases with an increase in temperature at the rate of about 1 per cent per degree C. at normal room temperatures. With selenium cells I is almost independent of temperature or may increase at the rate of about 0·2 per cent per degree C. The smaller variation in the latter case may be due to the relatively high resistance of the selenium cell compared with that of the copper-oxide type. For example, a selenium cell may have a resistance of 50,000 ohms per cm.^{2} against 5000 ohms per cm.^{2} for the copper-oxide cell.

The curves of Figs. 18–22 and 18–23 only refer to the static characteristics of photo-voltaic cells. As stated in dealing with metal rectifiers, such cells possess electrostatic capacity and this is indicated by the condenser in Fig. 18–20. As the reactance of this capacity decreases with increasing frequency, it is evident that it will have a shunting effect on R, thus decreasing the output current as the frequency of the incident radiation increases. Because of this, photo-voltaic cells are not employed with modulated light, but are chiefly used for photometry and relay operation. However, apart from the effects of capacity the time lag of a photo-voltaic cell is no larger than that of the photo-emissive type.

Photo-conducting Cells

In some respects the principle upon which the photo-conducting cell operates is similar to that of the rectifier cell. If radiation falls upon a semi-conductor, and the quantum energy is adequate, electrons will be raised from fully occupied or impurity levels to unoccupied, or running, levels. This condition of affairs, of course, tends to occur with all substances, but in good conductors the effect is imperceptible as normally large numbers of electrons already exist in running levels. In a poor conductor, however, an increase in the number of electrons in running levels is perceptible, and the immediate effect of the incident radiation is an increase in the conductivity of the substance. This condition is known as the inner photo-electric effect. The difference now between the rectifier and photo-conducting cells is one of boundary conditions. In the first case the energy of the electrons is such that they can surmount the potential barrier between the semi- and good conductors and thus produce an e.m.f. In the conducting cell it is necessary to apply an external e.m.f. to cause the electrons in the running levels to cross the potential barrier and so provide an external current. The magnitude of this current, of course, depends on the semi-conductor resistance and so on the intensity of the incident radiation.

CONSTRUCTION

The construction of a conductivity cell follows that shown by Fig. 18–25. A glass plate has two grid-like structures formed on its surface by depositing a film of gold thereon and then cutting the film to produce two separate electrodes. A thin film of semi-conductor is then deposited over the glass and grids and the whole finally enclosed in an evacuated glass vessel to avoid atmospheric contamination. The conductivity of the cell is, of course, that between the two grid-like electrodes. The semi-conductor is usually selenium, but alloys of selenium and tellurium are employed as well as thallous oxysulphide. The latter material is sensitive to radiation of longer wavelength than selenium, as shown by Fig. 18–28. As the electrical properties of a semi-conductor depend on impurities in its construction, it is evident that the performance of

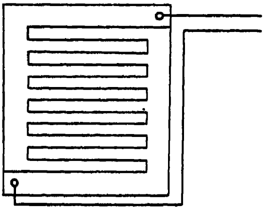


FIG. 18-25

a cell will depend on its preparation. It is also to be anticipated that considerable variations may be found between different cells.

CELL RESISTANCE

Conductivity cells are usually designed for use at 100 to 250 volts and, in the case of selenium, may have a resistance between 5 and 100 megohms. Thus the "dark" current lies between 1 and 50 μA. Ohm's Law is not exactly followed, even when the cells are unilluminated. After applying a voltage to a cell the current tends to increase, although at a diminishing rate. If a change of voltage occurs, the current increases more rapidly than does the voltage. Hence it follows that it is desirable that the voltage applied to a conductivity cell should not be subject to variations. As the resistance of a semi-conductor varies inversely with temperature, it is to be expected that conductivity cells are susceptible to temperature changes. This is so, the dark current increasing by about 2 per cent per 1° C. rise in temperature at normal room temperatures. When the cell is illuminated, the current still increases with temperature, but less rapidly than when the cell is dark. A further effect of temperature is to decrease the sensitivity of the cell as the temperature rises.

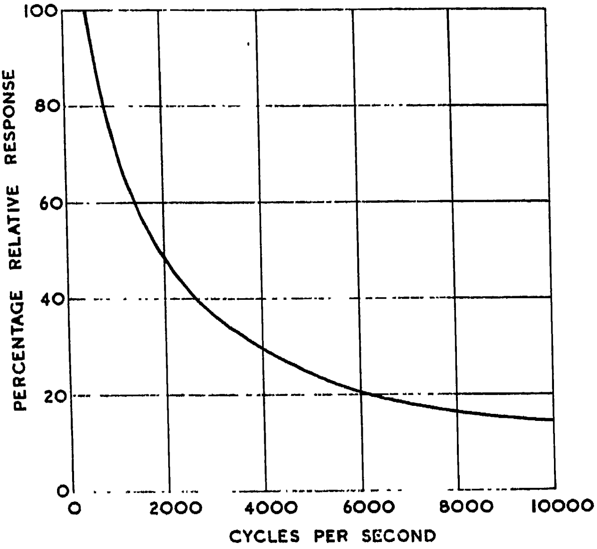


FIG. 18 26

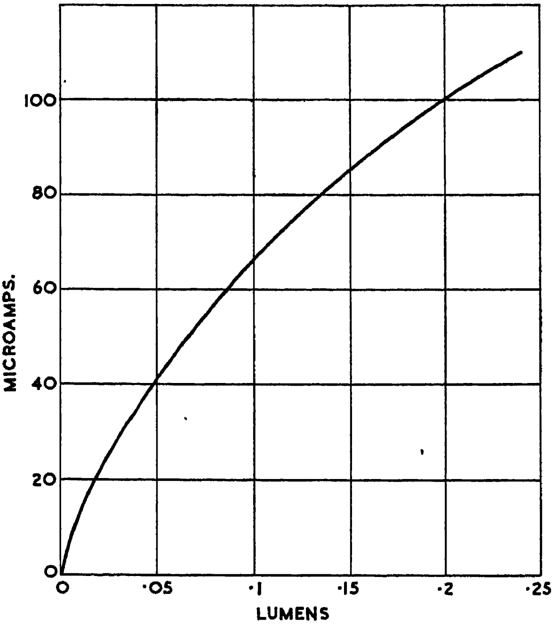


Fig. 18–27

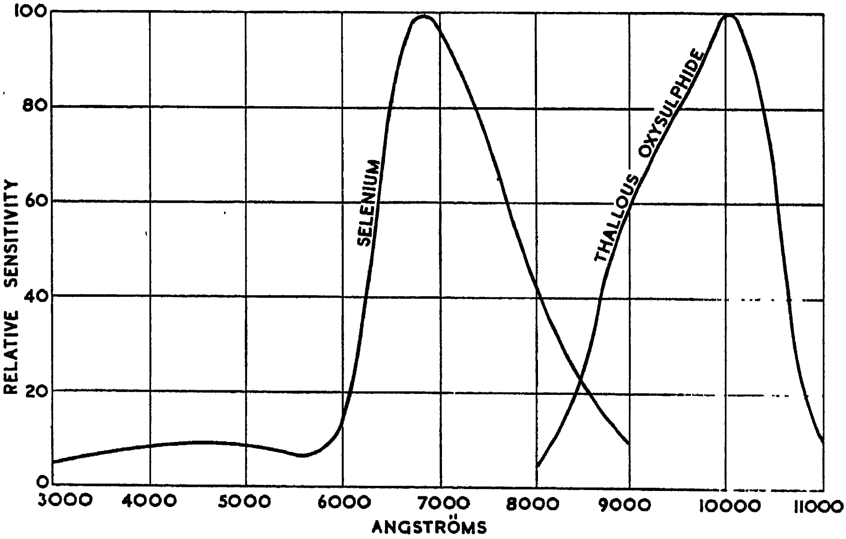


Fig. 18–28

Time Lag

It has been previously stated that when a voltage is applied to a conductivity cell the current slowly increases with time. A similar effect occurs when illumination is applied to the cell, the current increasing asymptotically with time. Similarly the current decreases asymptotically when the illumination is cut off. In early cells the time for the current to reach a practically steady value was measured in minutes, but improvement has been such that the time has been reduced to about one-tenth of a second. However, this latter time period indicates that the cell has a very poor frequency response and is almost useless for work in, say, the audio-frequency range. This is indicated by Fig. 18–26, which gives the relative frequency response of a selenium cell within the audio range. Because of these characteristics, photo-conductive cells are invariably employed with steady illumination only.

Sensitivity and Spectral Response

The sensitivity of a selenium cell when illuminated by a tungsten filament lamp is shown by Fig. 18–27. Unlike photo-emissive and photo-voltaic cells, the sensitivity or photo-electric efficiency is not constant but decreases with increasing illumination. Spectral response curves are given by Fig. 18–28, from which it will be noted that the cells are particularly responsive towards the red end of the spectrum.

CHAPTER XIX

ELECTRIC DISCHARGE LAMPS

Among the applications of the gaseous discharge, one that is fast becoming of outstanding importance is the employment of glows and arcs as light sources. The principal reason for this is that a gaseous discharge functions as a "cold" light source (in contrast with the usual temperature radiator) and in consequence operates at a higher efficiency than an incandescent lamp.

Temperature Radiation

As is commonly known, illumination is generally effected by heating of solids to incandescence. As shown in Chapter I, the energy radiation from such solids is spread over a theoretically infinite number of wavelengths, the energy distribution being governed by Planck's Law. However, from the viewpoint of illumination, we are chiefly concerned with an exceedingly narrow band of waves, namely those lying between 4000 Å to 7000 Å, for to these we owe the sensation known as light.

EFFICIENCY OF TEMPERATURE RADIATORS

The luminous efficiency of a temperature radiator can be expressed as

$$\frac{\text{Energy in Visible Spectrum}}{\text{Total Energy Radiated}} \quad . \quad . \quad (19\text{–}1)$$

but this does not take account of the varying sensibility of the eye in different parts of the spectrum. A given amount of energy is differently evaluated by the eye according to the wavelength of the energy, the eye being most sensitive at a wavelength of 5550 Å. The relative visibility curve for the average human eye is shown by Fig. 19–1, any ordinate being known as the relative visibility factor for the particular wavelength at which it is taken.

The rate of emission of luminous energy is measured in lumens, a source of one standard candle giving 4π lumens. If one watt could be entirely converted into luminous power at 5550 Å, then the luminous efficiency would be 625 lumens per watt. It follows that the efficiency at any other wavelength is obtained by multiplying 625 by the relative visibility factor. The efficiency of all light sources so far available falls far below the optimum figure quoted

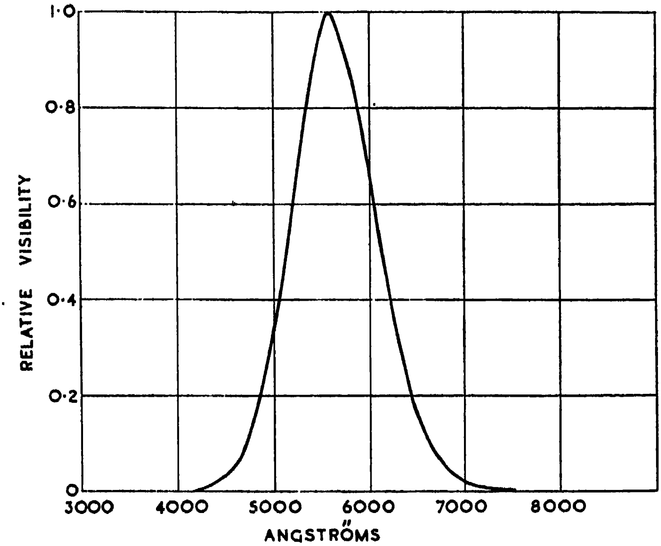


FIG. 19-1

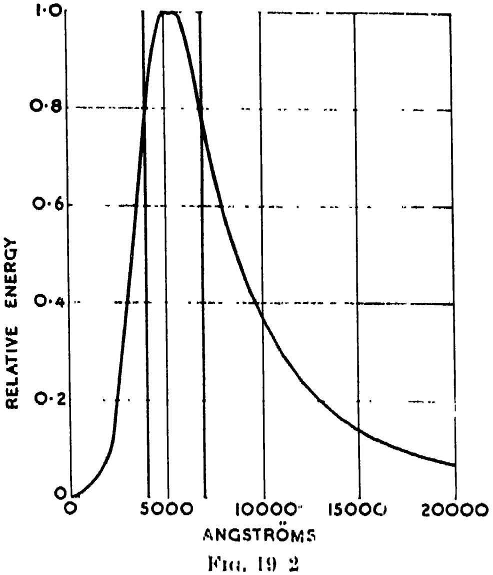


FIG. 19 2

above. For example, if it were possible to operate an incandescent lamp at 6000° K. so that the emission curve peaks in the visible spectrum, as shown by Fig. 19–2, the luminous efficiency would be approximately 86 lumens per watt. This is about six times as great as the present efficiency of such lamps. If all the energy were radiated within the visible spectrum, without losses in either the ultra-violet or infra-red, the efficiency would be approximately 225 lumens per watt. Except for lamps designed to have a high efficiency at the expense of a short life, it is not possible to run present materials as temperature radiators beyond about 2500° C. Because of this, incandescent light sources have luminous efficiencies of no more than about 15 lumens per watt.

If I_λ denotes the energy of radiation per square centimetre per second of a given radiator for wavelength λ, E_u the rate of energy radiation per unit area between the wavelengths λ_1 and λ_2 which define the limits of the visible region, E the total rate of energy radiation, and L the light emitted in lumens per unit area, then, from Fig. 19–2,

$$E = \int_0^\infty I_\lambda d\lambda \quad . \quad . \quad . \quad . \quad . \quad (19\text{–}2)$$

$$E_u = \int_{\lambda_1}^{\lambda_2} I_\lambda d\lambda \quad . \quad . \quad . \quad . \quad . \quad (19\text{–}3)$$

Hence (19–1) may be expressed as

$$\frac{E_u}{E} = \frac{\int_{\lambda_1}^{\lambda_2} I_\lambda d\lambda}{\int_0^\infty I_\lambda d\lambda}$$

which may now be termed the radiant efficiency.

If V_λ is the relative visibility factor for light of wavelength λ, then

$$L = 625 \int_{\lambda_1}^{\lambda_2} V_\lambda I_\lambda d\lambda \quad . \quad . \quad . \quad . \quad (19\text{–}4)$$

If η is the luminous efficiency, then

$$\eta = \frac{L}{E} = \frac{625\int_{\lambda_1}^{\lambda_2} V_\lambda I_\lambda d\lambda}{\int_0^\infty I_\lambda d\lambda} \quad . \quad . \quad . \quad (19\text{–}5)$$

The optimum luminous efficiency, η_0, is defined by

$$\eta_0 = \frac{L}{E_u} = \frac{625\int_{\lambda_1}^{\lambda_2} V_\lambda I_\lambda d\lambda}{\int_{\lambda_1}^{\lambda_2} I_\lambda d\lambda}$$

η_0 having already been stated to be approximately 225 lumens per watt. Actually η_0 represents the luminous efficiency of a fictitious source which has the same energy distribution as a given actual source, but which emits no energy outside the visible region. Some radiation characteristics for a black-body radiator are given by Table 19–1.

TABLE 19–1

TEMPERATURE ° K.	WATTS PER CM.^{2}	L LUMENS PER CM.^{2}	η LUMENS PER WATT	η_0 LUMENS PER WATT
2,000	91·4	$1·4 \times 10^2$	1·53	153
3,000	426·4	$8·89 \times 10^3$	20·8	206·7
4,000	1,462	$7·32 \times 10^4$	50·1	219·6
5,000	3,550	$2·64 \times 10^5$	73·9	229·3
5,500	5,226	4.22×10^5	82·6	227·9
6,000	7,403	$6·32 \times 10^5$	84·0	221·1
6,500	10,190	$8·75 \times 10^5$	86·3	218·5
7,000	13,170	$11·53 \times 10^5$	84·1	212·9

Luminescent Radiators

Because of the low efficiency of temperature radiators, others have been sought which do not depend on the temperature of a body. A number of such radiators have already been described in Chapter VII, most of these functioning by the effect of radiation or an electron beam on some solid substance. However, when an electrical discharge occurs through a gas this may, as described in Chapter II, produce luminous radiation in the form of a glow or an arc. This form of radiation is termed electro-luminescence and is due to energy exchanges within the atom in accordance with Bohr's theory. Compared with temperature radiators which produce a continuous spectrum (i.e. where all frequencies are present), an electro-luminescent radiator radiates over relatively few wavelengths and produces a line spectrum. These two forms are contrasted by Fig. 19–3. The spectrum of the electrical discharge is such that possibilities exist of producing a relatively high proportion of radiation within the visible spectrum without the wasteful infra-red. Some heat, however, is generally essential to maintain the discharge, and frequently a considerable amount of energy is radiated in the ultra-violet. The latter may, however, be converted into visible radiation in the manner described in Chapter VII, thus further increasing the efficiency of the discharge lamp.

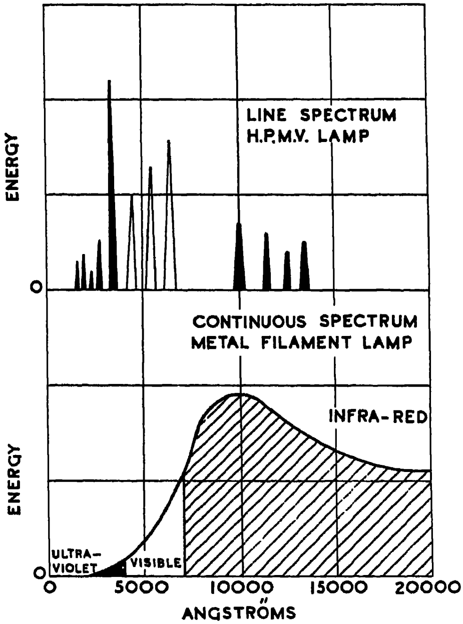


FIG. 19–3

Low-pressure, Cold-cathode Discharge Lamps

The degree of luminosity accompanying a discharge varies with the gas employed, certain gases being quite useless as illuminants. Among the common gases showing a useful degree of illumination are carbon dioxide and nitrogen, both of these having a relatively high efficiency of about two watts per candle-power. Apart from illuminating properties and low efficiencies, various other factors prevent the majority of gases from being employed in practical discharge tubes. The most notable of these is the gradual lowering of the pressure of the gas-filling. The effect of this is to necessitate a continually increasing voltage to strike the tube until the value of the required voltage becomes quite impractical. The cause of this is twofold, namely, chemical action at the electrodes and a phenomenon known as sputtering or splashing. As an instance of the first kind, if the gas-filling is carbon monoxide and the electrodes are iron, the gas will combine with the electrodes to form oxide of iron. Gases which are almost inert under normal conditions may become chemically active when ionized, resulting in similar effects. The time for such effects to occur is relatively short and, in the case of magnesium or aluminium electrodes, with a filling of oxygen, is only a few seconds.

By employing an inert gas in discharge tubes or electrodes inert to the gas-filling, chemical effects as described above cannot occur. However, even in these circumstances the gas may gradually disappear, due to sputtering. This is due to positive-ion bombardment of the cathode resulting in the ejection of particles from the latter. Such particles envelop the gas molecules and are deposited on the walls of the discharge tube. Examination of a gaseous discharge tube which has been in use for a prolonged period will show a darkening of the glass due to the foregoing phenomenon. The gradual disappearance of the gas is sometimes referred to as "clean-up," and the condition of the tube as "hardening."

THE RARE GASES

The discovery in 1895 of what are known as the rare gases opened up considerable possibilities in discharge lighting. These gases (argon, neon, helium, krypton, and zenon) were found to exist in the atmosphere, but had hitherto escaped detection because of the relatively small quantities present. Hence the term "rare gases." An extremely important property of this group of gases is their chemical inertness or their inability to form chemical combinations with any other element. Thus, the employment of such gases in discharge tubes should, and does, prevent clean-up and hardening due to chemical actions at the electrodes. However, only neon and helium are employed, because of the scarcity and poor luminosity of the other three.

THE CATHODE FALL

The lowest cathode fall is furnished by the alkaline metals, such as sodium, potassium, and rubidium, while next above these come the earth alkaline metals such as calcium, barium, and magnesium. Iron, silver, copper, and zinc give a higher cathode fall than the foregoing group, the highest fall for metals being obtained with mercury, gold, and platinum. Table 19–2 shows the cathode fall, in volts, for various electrodes in different gases.

TABLE 19–2

CATHODE	Pt	Fe	Al	Mg	Na	K
Neon . .	165	153	120	94	75	68
Helium . .	165	161	164	125	80	69
Hydrogen . .	300	290	190	168	185	172
Nitrogen . .	232	222	210	207	178	170

Although from the foregoing table it appears most advantageous to employ the alkaline metals for electrodes, this is not so for the following reasons. These metals rapidly oxidize when exposed to air, they would be attacked by foreign gases present in a tube during exhaustion, and, furthermore, they are expensive. Thus it is evident that for electrode materials, the choice must fall on either copper, aluminium, or iron, and in practice, this is the case, the last two metals being commonly employed. Aluminium, however, cannot be used in mercury tubes, because it is rapidly attacked by mercury.

Further factors affecting sputtering are the electrode current density and gas pressure. As discussed in Chapter II, if the cathode glow is equal in cross-section to that of the cathode, any increase in current will result in an abnormal cathode fall. This usually occurs in luminous discharge tubes, although the diameter of such tubes is usually increased to accommodate electrodes of greater diameter than that part of the tube in which is situated the positive column. A result of the abnormal cathode fall is larger cathode voltages than shown by Table 19–2. Some typical figures for the principal gas-fillings are given by Table 19–3 at top of page 627.

The pressure of the gas-filling affects the electrical conductivity and luminous efficiency of a discharge tube as well as sputtering. Regarding the latter, the lower the gas pressure the greater the velocity with which ions strike the cathode. This leads to an increase in sputtering, and from this viewpoint the gas pressure should not

TABLE 19–3

	5 mA per Cm.^{2}	2 mA per Cm.^{2}	0·7 mA per Cm.^{2}
Neon . .	425	345	295
Mercury . .	300	260	245
Helium . .	575	525	325

be too low. From the viewpoint of electrical conductivity and luminous efficiency the optimum pressure for neon tubes is about 1 mm. Hg. However, this pressure is too low to prevent sputtering, and in practice the pressure of the gas filling is between 5 and 10 mm. Hg. In these circumstances the luminous efficiency is about 15 lumens per watt.

The Positive Column

The light source of "neon" lamps is, of course, the positive column. When the filling is helium the colour of the discharge is creamy white, mercury vapour gives a blue discharge, and neon, of course, the familiar red. Energy level diagrams for neon and mercury are shown by Figs. 19–4 and 19–5, and will explain the characteristic colours for lamps with these fillings. It will be noted that near ultra-violet radiation occurs with mercury and far ultra-violet with neon. As explained in Chapter VII, this radiation may be turned to useful account by causing it to excite fluorescence in certain materials.

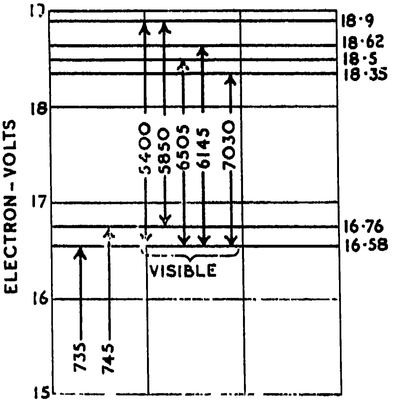


Fig. 19–4

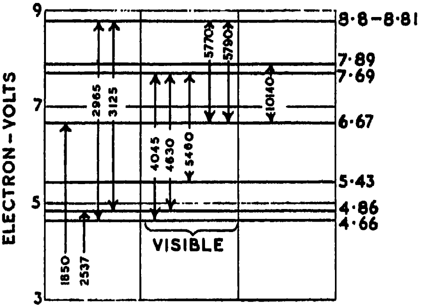


Fig. 19–5

The voltage drop along the positive column is proportional to its length and decreases with an increase in tube diameter or current. The diameters commonly employed are 10, 15, and 20 mm. and the volt drop per foot for the main gas-fillings is shown by Table 19–4.

TABLE 19–4

Gas	25 mA in 10 mm. Tube	35 mA in 15 mm. Tube	75 mA in 20 mm. Tube
Neon	175	120	90
Mercury	130	95	70
Helium	370	260	190

These figures must be taken to represent only average values, as the actual volt drop depends on the gas pressure and purity and the manufacturing process. However, Tables 19–3 and 19–4 may act as a guide when first computing the necessary operating voltage of an installation. It must be remembered that the figures given by the tables refer to the running voltage. The initial value of the striking potential is generally larger than that under running conditions and hence allowance must be made for this. Thus with tubes of small diameter, the starting voltage may be as much as 100 per cent greater than the running voltage, while large diameter tubes may necessitate a starting voltage 60 per cent greater than the running voltage.*

Transformer Operation of Neon Lamps

Tables 19–3 and 19–4 indicate that voltages considerably above those normally available are necessary for operating neon tube installations. This requirement is simply met by means of a step-up transformer. The installation must, of course, be "controlled," and this may be effected by placing the necessary impedance in series with either the primary or secondary circuit. For design and calculation purposes, a primary impedance may be replaced by an equivalent secondary one, the value of the latter being given by the primary impedance multiplied by the square of the transformation ratio. However, general practice with neon installations is to employ a leaky, or stray field, transformer. The advantage of this is that the necessary impedance can be incorporated in the transformer design, this leading to a simpler and cheaper installation than where separate impedances are employed.

---

* As an example, it may be stated that for the lamp of Fig. 19–8, the initial striking voltage is 1250 volts, while the running voltage is 600 volts.

Operating Characteristics of Neon Discharge Tubes

Because of the negative resistance characteristic of a glow discharge it is not possible to operate discharge tubes directly from the supply voltage. As already stated, this means that an impedance must be included in series with the tube. The impedance may take the form of a condenser or resistance, although a choke is most commonly employed. Assuming a series resistance is used, the voltage across the lamp increases until the striking voltage V_s is reached. At this stage current commences to flow, the value of which is given by

$$i = \frac{V_m \sin pt - V_L}{R} \quad . \quad . \quad . \quad (19\text{–}6)$$

where $V_m \sin pt$ is the instantaneous value of the supply voltage, V_L the lamp voltage, and R is the value of the series resistance.

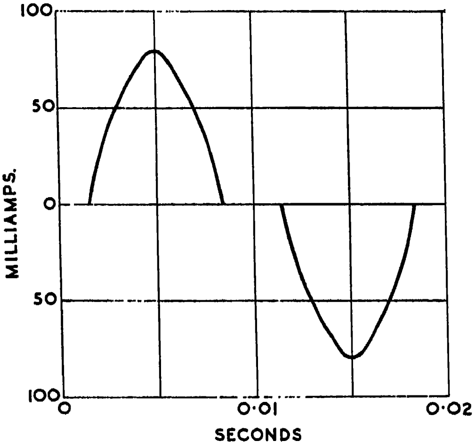


FIG. 19–6

Now V_L is composed of the cathode fall and the drop in the positive column. As previously stated, an abnormal cathode fall occurs with neon lamps, which means that an increase in lamp current will cause an increase in cathode fall. The effect of an increase in current, however, is to decrease the drop in the positive column, and, in practice, it is found that these two effects tend to counterbalance each other. This results in the lamp volt drop being independent of current and hence V_L may be regarded as approximately constant during current flow.

From the foregoing and (19–6) the current waveform may be deduced, a typical case being shown by Fig. 19–6, where $V_m = 700$, $V_L = 300$, and $R = 5000$ ohms. It is evident that conduction does not commence until $V_m \sin pt = V_s$ or $pt = \text{arc sin } V_s/V_m$. In many cases the lamp strikes and runs at approximately the same voltage, in which case $pt = \text{arc sin } V_L/V_m$. The lamp extinguishes at a voltage roughly equal to V_L and hence extinction occurs for V_m sin $pt = V_L$ or $pt = \text{arc sin } (\pi - V_L/V_m)$.

It is evident that with resistive control the current does not persist for a full half-cycle. Actually the fraction of a cycle during which conduction occurs is

$$(\pi - 2V_L/V_m)/\pi$$

Several disadvantages are attached to this. Firstly the stroboscopic effect experienced with discharge lamps is enhanced by a short conduction period. Secondly, the value of the striking voltage in each half-cycle is affected by the length of the period the lamp is extinguished. When the lamp current reaches the zero value, a certain number of ions are still present in the discharge tube, and these commence to recombine. Providing a sufficient voltage is available to cause the tube to restrike, relatively little recombination occurs and the restriking voltage is less than that necessary to strike the tube initially, i.e. when the supply switch is first closed. Reference to Fig. 19–6 shows that when extinction occurs a considerable interval follows before restriking takes place. During this interval de-ionization of the tube is occurring necessitating a relatively high restriking voltage. It is evident that the higher the restriking voltage the shorter the conduction period, for actually the fraction that the latter is of a cycle is given by

$$\left(\pi - \frac{V_L + V_s}{V}\right)\bigg/\pi$$

rather than by $$(\pi - 2V_L/V_m)/\pi$$

Naturally the reduced conduction period results in increased stroboscopic effects.

Because of the foregoing effects and the losses which a resistance incurs, this form of control is only employed with the smallest types of lamps. Condenser control obviates the losses, but suffers from the same disadvantage as resistive control with regard to a short conduction period.

INDUCTIVE CONTROL

The usual method of controlling discharge lamps is by means of a series-connected choke. The latter has relatively small losses and possesses the advantage of causing a longer conduction period than can be obtained with either a resistance or condenser. In fact, with inductive control, the current wave is usually spread over the complete half-cycle as shown by the oscillogram of Fig. 19–7. Greater details of this will be given later, but it may be stated that the improvement in the current waveform is due to the self-induced voltage of the choke giving a further source of potential to the lamp when the supply potential falls below V_L.

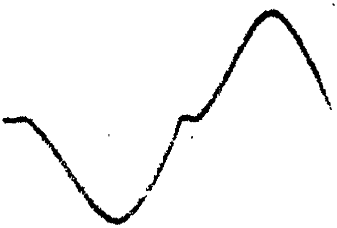


FIG. 19–7

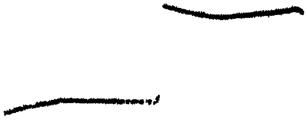


FIG. 19–8

VOLTAGE AND CURRENT WAVEFORMS

The oscillogram of Fig. 19–7 shows the current waveform of a neon tube, 43 cm. in length and 15 mm. in diameter, when operated in series with a 21-henry choke. From this it will be noted that the current persists for almost the full half-cycle. Fig. 19–8 shows the almost rectangular waveform of the lamp voltage, while Fig. 19–9 is an oscillogram of the voltage across the choke. An explanation of the somewhat peculiar character of the latter will be given later. It is evident from Fig. 19–8 that the striking voltage

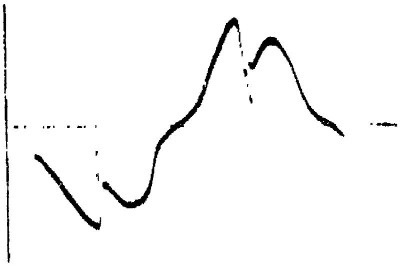


FIG. 19–9

is somewhat in excess of the running voltage; i.e. $V_s > V_L$.

High-pressure Mercury-vapour Lamps

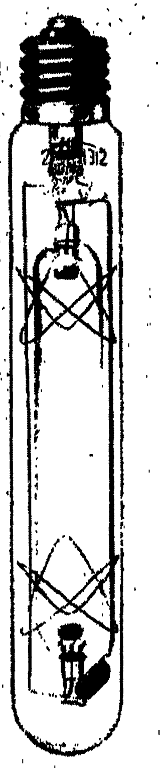
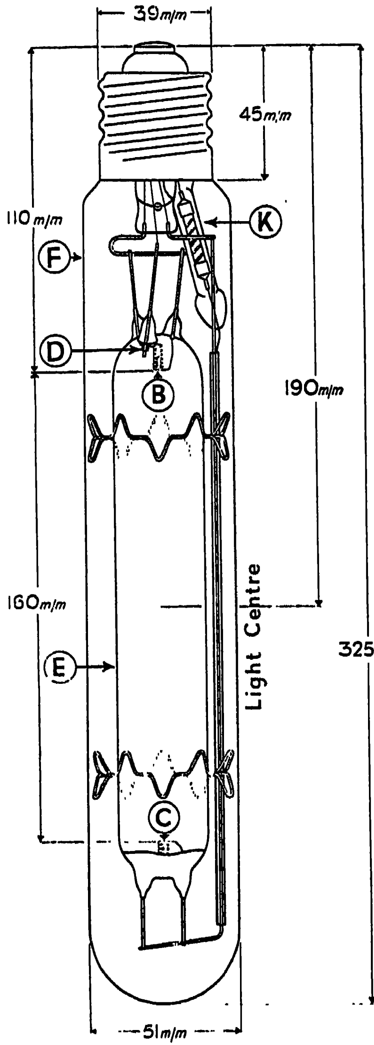


FIG. 19–10

By courtesy of Siemens Electric Lamps and Supplies Ltd.

The principal application of the lamps so far described is to decorative and advertising purposes rather than to the illumination of surface areas. By employing mercury vapour at high pressures (an atmosphere and above) lamps having much higher efficiencies can be produced suitable for floodlighting large areas. A typical 400-watt h.p.m.v. lamp is shown by Fig. 19–10. It consists of a discharge tube enclosed within an envelope of ordinary glass. The tube is formed of special hard glass or quartz, according to the temperature which must be withstood. At the ends of the inner tube are electrodes which consist of activated rare-earth oxides held within a tungsten spiral. During the passage of the current these electrodes become heated and emit electrons by the process of thermionic emission. At the upper end of the discharge tube is an auxiliary electrode placed close to the main electrode. This serves the purpose of initiating the discharge. The discharge tube contains a rare gas, such as argon, at a pressure of about one hundredth of an atmosphere, with a carefully measured quantity of mercury. The space between the tube and envelope is filled with oxygen at a low pressure.

In order to understand the functioning of the lamp, reference may be made to Fig. 19–11. On connecting the lamp (via a choke) to the supply voltage, the gas pressure is too low, and the main electrode spacing too large, to permit a discharge to occur between the main electrodes. It will be noted, however, that the full potential difference is developed between the top main and auxiliary electrodes. Due to their proximity a glow discharge occurs, the glow current being limited to a few milliamperes by the resistance K. The glow discharge initially occurs in the argon due to the very low vapour pressure of mercury at normal temperature. The effect of the glow discharge is to provide sufficient ions for the main discharge to start. Once the main discharge is struck, the heat due to this commences to vaporize the mercury, thus raising the mercury-vapour pressure. The bombardment of the electrodes by mercury-vapour ions also raises the temperature of the former, which then become electron emitters. At the commencement of the discharge the tube is filled with light, but the luminosity is low. As the temperature rises the luminous column becomes narrower and brighter, full brilliancy being attained after about five minutes. The function of the outer envelope is to conserve the heat of the discharge tube in order to assist in the vaporization of the mercury and maintain as uniform a temperature as possible. It also absorbs any harmful ultra-violet radiation. The ends of the discharge tube are silvered to reduce heat radiation from the electrodes.



B & C: Main Electrodes
K: 50000-ohm Resistance
D: Auxiliary Electrode
E: Discharge Tube
F: Glass Envelope

FIG. 19-11

By courtesy of the Edison Swan Electric Co., Ltd.

PHOTOMETRIC AND OPTICAL CHARACTERISTICS

The light source of the lamp is the narrow cord-like discharge between the electrodes, the average brightness of this being approximately 120 candles per cm.^{2} It is at once evident that, compared with a tungsten filament lamp (500 candles per cm.^{2}), this is low, and

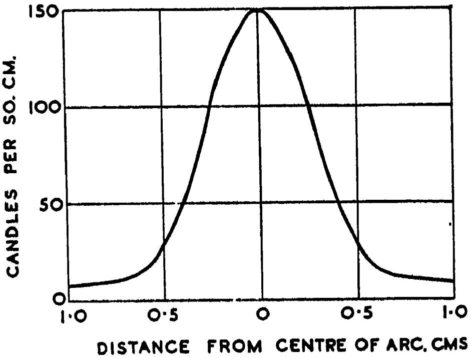


FIG. 19-12

thus h.p.m.v. lamps are relatively free from glare. The boundary of the luminous column is not clearly defined and, of course, the brightness is not uniform across its section. Fig. 19–12 shows a brightness distribution curve for a 400-watt lamp, and taking the edge of the column to be where the brightness is one-tenth of the maximum, the width of the column is 1·2 cm.

A light distribution curve for a 400-watt lamp is shown by the polar diagram of Fig. 19–13, this curve being taken in a vertical plane which contains the lamp axis. The total light output is 18,000 lumens, the efficiency being, therefore, 43 lumens per watt.

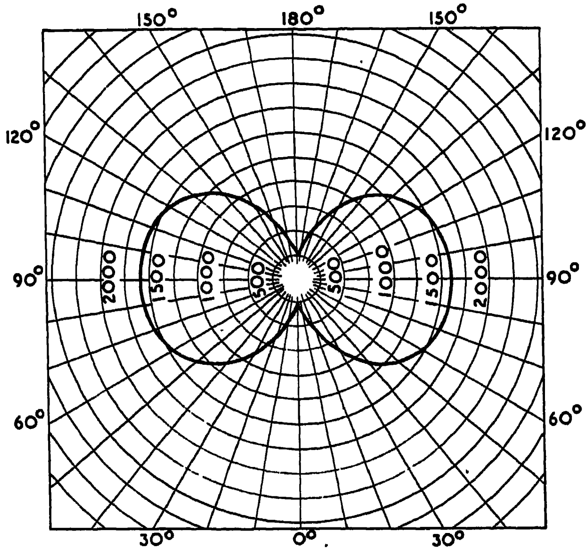


FIG. 19–13

OPTICAL CHARACTERISTICS

The majority of the light from mercury-vapour lamps, whether of the high-pressure or low-pressure types, consists of line spectrum radiation. This is due to the atoms within the positive column being excited to certain energy levels with, of course, the object of producing light of a particular colour. As shown in Chapter II, within the positive column the field gradient is relatively low and high-excitation levels are produced by cumulative electron impacts. The field gradient depends on the gas or vapour density, increasing with an increase in these. Thus the current and vapour densities will depend on whether high- or low-energy level excitation is desired. Referring to Fig. 19–14, it will be noted that the 4·86 and 6·67-volt states are in the ultra-violet region, and hence higher levels of excitation are required as indicated by the diagram. Atoms excited to these higher states, such as the 7·69 volt, return to the ground state via the 4·86- and 4·66-volt states. Hence visible and invisible radiation are produced, and the higher the ratio of the former to the latter the greater is the efficiency of the lamp. Thus, the efficiency of the lamp increases with an increase in vapour and current density, the maximum value occurring for a pressure of about 50 atmospheres and a loading of 100 watts per centimetre length of arc. As the pressure and arc loading are increased, in addition to the normal line spectrum, a weak continuous spectrum is formed. This is due to the increasing temperature causing the mercury vapour to act as a temperature radiator.

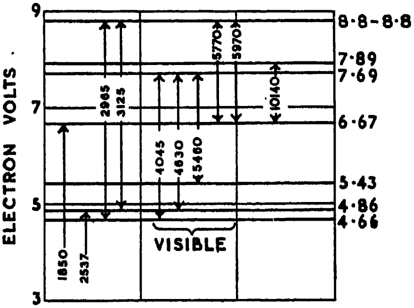


FIG. 19-14

PHYSICAL AND ELECTRICAL CHARACTERISTICS

To obtain stable operation of h.p.m.v. lamps they are designed to operate with all the mercury vaporized. In these circumstances variations in supply voltage and ambient temperature have but slight effect on the lamp voltage over wide limits. This voltage, V_L, is roughly constant during conduction, as indicated by Fig. 19–20, from which it will be noted that the lamp voltage waveform is roughly rectangular.

TEMPERATURE

In contrast with l.p.m.v. lamps those of the high-pressure type operate at relatively high temperatures. In the case of the 400-watt lamp, about 20 per cent of the energy supplied is radiated as light. A further 10 per cent is emitted as ultra-violet and near infra-red. The remaining 70 per cent is emitted as infra-red radiation, and has the effect of heating the discharge tube to between 500° C. and 600° C. The outer glass envelope attains a temperature of about 350° C.

An effect of the lamp temperature is its influence on the re-striking of the lamp should it be switched off. When this occurs the electrodes rapidly cool and thus become ineffective as electron emitters. The vapour pressure is too high to allow restriking of the lamp, and hence a period must elapse in which mercury vapour may condense and the pressure fall before restriking will occur. With a 400-watt lamp this is about three minutes in still air and for a 125-watt lamp two minutes.

ELECTRICAL CHARACTERISTICS

As h.p.m.v. lamps operate directly from normal supply voltages, no transformer is needed, but a series impedance is essential, as with cold-cathode low-pressure tubes. This generally takes the form of an iron-cored choke, although a resistance or condenser may be used. Of course, on direct current a resistance is essential. As previously stated, the voltage across the lamp, V_L, is roughly constant over a cycle. The value of this voltage is determined by the quantity of mercury within the lamp being higher the greater the quantity. During the time the lamp is running up, i.e. before the mercury is entirely vaporized, V_L increases with an increase in lamp watts. When the mercury is fully vaporized, the lamp voltage tends to remain constant with an increase in lamp watts, as indicated in Fig. 19–15. In determining the quantity of mercury to be used, and hence the voltage at which a lamp will operate, several factors must be considered. These are the change in lamp watts for a given change in supply voltage, the change in lamp

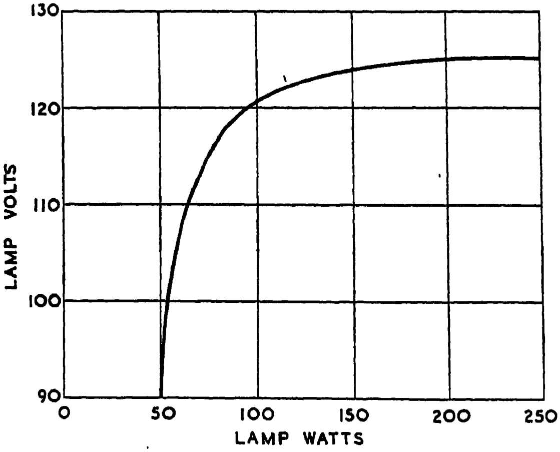


FIG. 19–15

watts for manufacturing variations in V_L, the power factor of the lamp-choke unit, and the permissible drop in the supply voltage for the lamp not to be extinguished. Table 19–5 sets out the effect of the ratio of V_L/V on these various factors. In practice the value adopted of V_L/V is about 0·6.

The Choke

Chokes employed for h.p.m.v. lamp control are iron-cored and possess a small air gap. The effect of the latter is to maintain the choke inductance approximately constant with changes in current and to permit adjustment (by means of varying the gap length) of the inductance between fine limits. It is, of course, desirable that the choke dimensions, losses, and cost shall be as small as possible. In practice the choke losses amount to about 5 per cent of the lamp wattage.

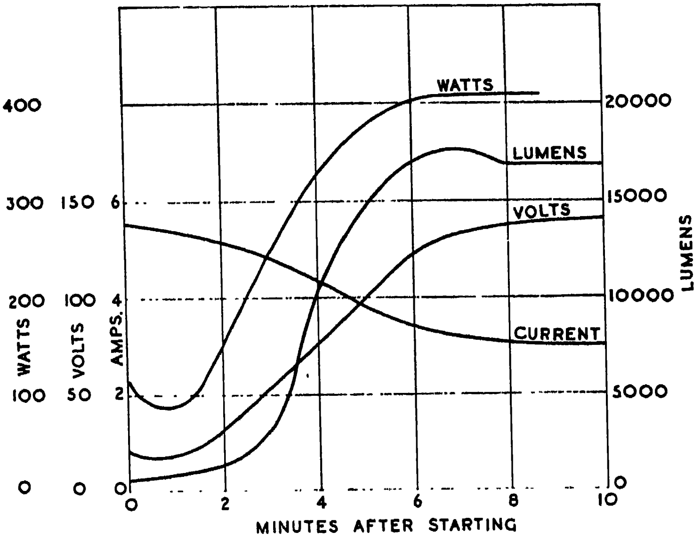
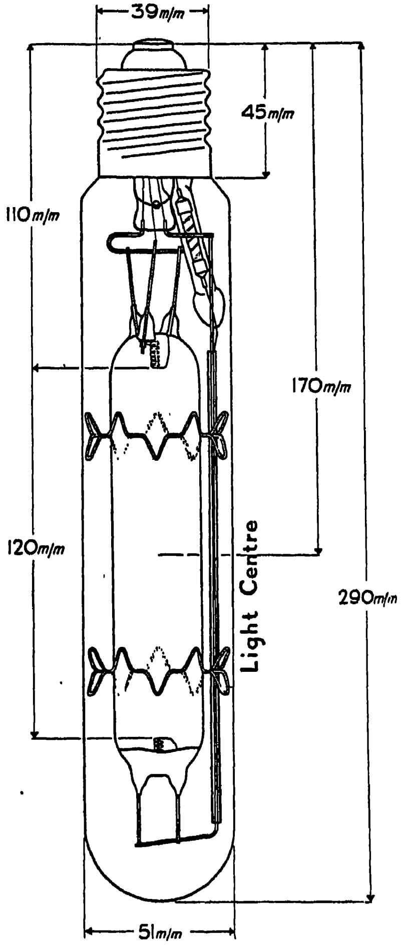


FIG. 19–16

Operating Characteristics

At the moment of closing the supply switch the full mains voltage is developed between the main and auxiliary electrodes, thus initiating the discharge. The discharge started, the lamp voltage falls to approximately 20 volts, with the result that, initially, practically the full supply voltage is across the choke. Hence the lamp current is a maximum almost immediately after closing the switch. Full lamp brilliancy is reached when the mercury-filling is completely vaporized. Fig. 19–16 shows a run-up curve for a 400-watt lamp.



250 Watt
ESCURA
LAMP

Fig. 19–17

By courtesy of the Edison Swan Electric Co., Ltd.

Some technical data for 400-watt and 250-watt lamps are given by Table 19–6.

The lamp sizes normally available are the 400-watt already described, 250-, 125-, and 80-watt. The 250-watt lamp is similar in structure to that of the 400-watt, and is shown by Fig. 19–17. The 125-watt and 80-watt types, however, differ somewhat from the two larger lamps. As stated previously, the luminous efficiency of h.p.m.v. lamps depends on the current and mercury-vapour densities, increasing with an increase in these quantities. The maximum efficiency is about 60 lumens per watt. Hence, if the dimensions of a discharge tube are constant, a reduction in wattage will result in a reduction in luminous efficiency. For reasons indicated by Table 19–5 the lamp voltage for most lamps is maintained between 100 and 140 volts. Thus, a reduction in current, and hence current density, reduces the arc loading, and, as already stated, the luminous efficiency.

In order to maintain a reasonably high efficiency with 125- and 80-watt lamps it was necessary to increase the arc loading and vapour pressure. As shown by Table 19–6, the arc loading of 250- and 400-watt lamps is 20·8 and

TABLE 19–5

V_L/V	Percentage Change in Lamp Watts for 1 per cent Change of V_L	Percentage Change in Lamp Watts for 1 per cent Change of V	Permissible Instantaneous Drop in V Without Extinction Occurring	Power Factor
0·1 0·2 0·3 0·4 0·5	1 per cent or less	2 per cent or less	Over 20 per cent of V	0·4 or less
0·6 0·7	1 per cent to 2 per cent	2 per cent approx.	10 per cent to 20 per cent of V	0·5 to 0·6
0·8 0·9	3 per cent to 10 per cent	3 per cent to 15 per cent	Less than 10 per cent of V	0·7 or over

TABLE 19–6

Lamp Watts	250	400
Supply volts	200–250 a.c.	200–250 a.c.
Lumens per watt	36	45
Maximum brightness of arc, candles/cm.^{2} .	150	150
Length, mm.	290	330
Arc length, mm.	120	160
Diameter, mm.	48	48

25 watts per cm. respectively. The figures for the 125- and 80-watt lamps are 42 and 40 watts per cm. respectively. These higher loadings naturally lead to an increase in brightness and to avoid glare the arc tubes of the smaller lamps are enclosed in pearl-glass bulbs.

The operating pressure of 80- and 125-watt lamps is between 5 and 10 atmospheres, while 250- and 400-watt lamps operate at approximately 1 atmosphere.

An effect of the higher arc loading is an increase in operating temperature which cannot be withstood by a glass-discharge tube. Because of this the discharge tube of the 80- and 125-watt lamps is made of quartz, which will withstand easily temperatures up to 1000° C. An increase in efficiency would be obtainable by the employment of quartz (and higher arc loadings) in 250- and 400-watt lamps, but the high cost of the material militates against its use in the larger lamps.

The constructional details of the 80-watt lamp are shown by Fig. 19–18, from which the relatively small dimensions of the arc tube will be noted. Technical data for this lamp as well as the 125-watt lamp are given by Table 19–7.

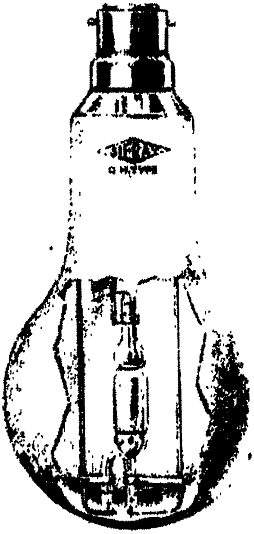


FIG. 19–18

By courtesy of Siemens Electric Lamps and Supplies, Ltd.

VOLTAGE AND CURRENT WAVEFORMS

Immediately after closing the supply switch of an inductively controlled h.p.m.v. lamp, practically the full supply voltage is developed across the choke. Hence the lamp current and choke voltage are, initially, approximately sinusoidal. As the mercury becomes vaporized the voltage across the lamp increases, the current, lamp, and choke voltages, under normal operating conditions, being shown by Fig. 19–19, 19–20, and 19–21 respectively. It will be noted that the lamp voltage is roughly rectangular and that the current persists for the full half-cycle. Occasionally h.p.m.v. lamps are controlled by a series resistance which may take the form of a glowing filament lamp. In these circumstances the current does not persist for 180° in each half-period, as it generally does with inductive control. Fig. 19–22 shows the current wave of a 125-watt lamp with resistive control when in the fully run-up condition. The voltage across the resistance is, of course, of the same form as the current wave, while the lamp voltage wave is shown by Fig.

TABLE 19–7

LAMP WATTS	80	125
Supply volts	200–250 a.c.	200–250 a.c.
Lumens per watt	38	40
Maximum brightness of arc, candles/cm.^{2} .	800	800
Maximum brightness of bulb, candles/cm.^{2} .	60	60
Length, mm.	160	178
Arc length, mm.	20	30
Diameter, mm.	80	90

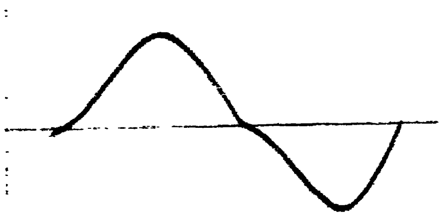


Fig 19–19

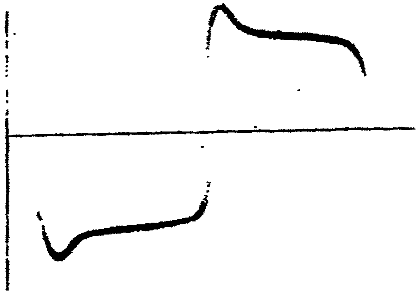


Fig. 19–20

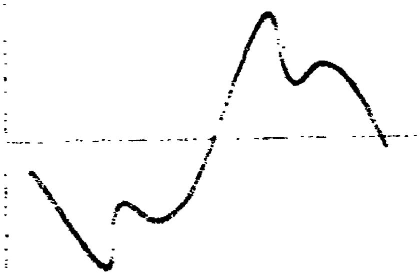


Fig. 19–21

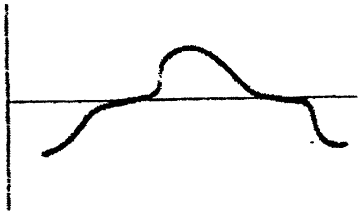


Fig. 19–22

19–23. Due to the distorted current wave, the power factor is less than unity and it is not possible to raise this in any way.

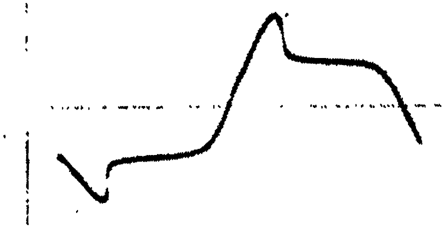


FIG. 19–23

The Sodium-vapour Lamp

The sodium-vapour is similar in principle to the h.p.m.v. lamp, but differs largely in its characteristics. A typical lamp is shown by Fig. 19–24, from which it will be noted that the discharge tube is U-shaped and contained in an outer tube. The purpose of shaping the discharge tube in this manner is to obtain the greater efficiency that accompanies long tubes; for increasing the length renders the energy dissipated at the electrodes a smaller fraction of the energy expended along the length of the tube, the latter energy, of course, being responsible for luminosity. Oxide cathodes at the ends of the tube act as arc-heated electron emitters in exactly the same sense as in the h.p.m.v. lamp.

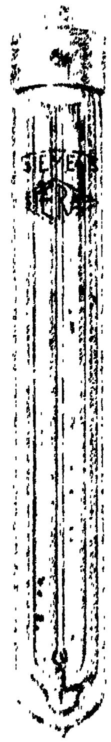


FIG. 19–24

By courtesy of Siemens Electric Lamps and Supplies, Ltd.

The discharge tube must be specially constructed to withstand the effect of hot sodium-vapour over long periods. This construction takes the form of a flashed or two-ply glass. The outer layer is ordinary soda glass, the inner layer consisting of a flashing of special glass designed to resist the effects of sodium. The discharge tube must be surrounded by a glass envelope in order to conserve the heat for the purpose of vaporizing the sodium. The envelope is in the form of a double-walled vacuum flask, from which the discharge tube may be detached when desired.

OPERATION

Distributed along the discharge tube are drops of sodium which are solid at normal temperatures. In consequence the sodium-vapour pressure is extremely low and the tube is filled with neon at a pressure of several millimetres of mercury for starting purposes. On connecting the lamp, via a suitable impedance, to the supply voltage the discharge first occurs in the neon, with the result that initially the appearance of the lamp is similar to that of a neon tube. As the time of operation and the temperature increase, the sodium melts, its vapour pressure increases, and the discharge tends to occurrather in the sodium vapour than in the neon. When the lamp is fully run-up the sodium is partially vaporized and the operating temperature is approximately 250° C. In these circumstances the lower excitation and ionization potentials of sodium (2·1 volts and 5·1 volts, respectively) cause the discharge to occur in this rather than in the neon.

The lamp is made in four sizes, 45-, 60-, 85-, and 140-watts respectively. The light from sodium lamps is practically mono-chromatic, being radiated by two yellow lines in the spectrum at 5890 Å and 5896 Å. Due to the absence of ultra-violet radiation, and the light occurring in that part of the spectrum to which the eye is most sensitive, the luminous efficiency of the sodium lamp is considerably higher than that of the h.p.m.v. lamp. Figures for the 45-, 85-, and 140-watt lamps are, respectively, 55·5, 71·1, and 71·5 lumens per watt.

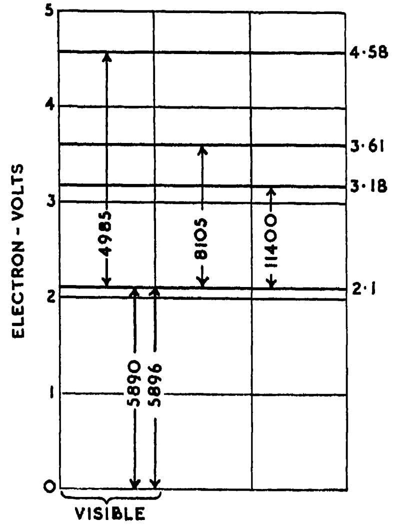


FIG. 19·25

OPTICAL CHARACTERISTICS

As stated above, the radiation of the sodium lamp, similarly to the h.p.m.v. lamp, consists of line spectrum radiation. Reference to the energy level diagram of Fig. 19–25 for sodium indicates that light results from the return from the 2·1 volt state to the ground state. Thus, in contrast to the h.p.m.v. lamp, the sodium lamp must be excited to give low energy level transitions. In practice this means that the arc loading must be relatively small, this being effected by low vapour pressures and current densities. For the 45-, 60-, 85-, and 140-watt lamps the arc loadings are 1·8, 1·95, 2·0, and 2·6 watts per cm., respectively.

The operating temperature of the sodium lamp is somewhat critical, the maximum efficiency occurring at between 220° C. and

250° C. However, due to the low operating pressure, temperature has no effect on the restriking characteristics; i.e. the lamp will immediately restrike if extinguished and then switched on again.

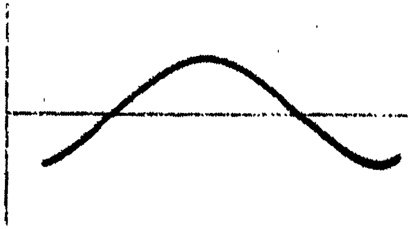


FIG. 19–26

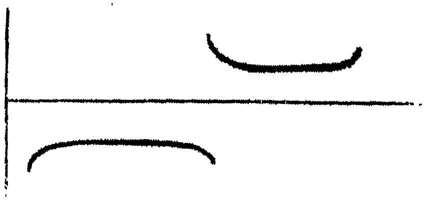


FIG. 19–27

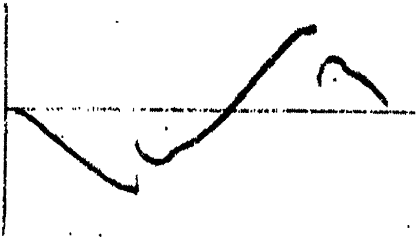


FIG. 19–28

VOLTAGE AND CURRENT WAVEFORMS

The current waveform of an inductively controlled 60-watt sodium-vapour lamp is shown by Fig. 19–26. The lamp-voltage waveform is shown by Fig. 19–27. Following a somewhat high striking potential of about 200 volts, the lamp voltage rapidly falls to approximately 100 volts, at which value it tends to remain during each half-cycle. The choke-voltage waveform is shown by Fig. 19–28.

ELECTRICAL CHARACTERISTICS

As previously stated, the sodium lamp has a gas-filling of neon to facilitate starting. No starting electrode is employed, and hence the momentary starting potential is somewhat high, about 440 volts.

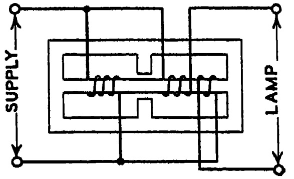


Fig. 19–29

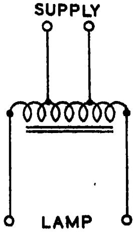


Fig. 19–30

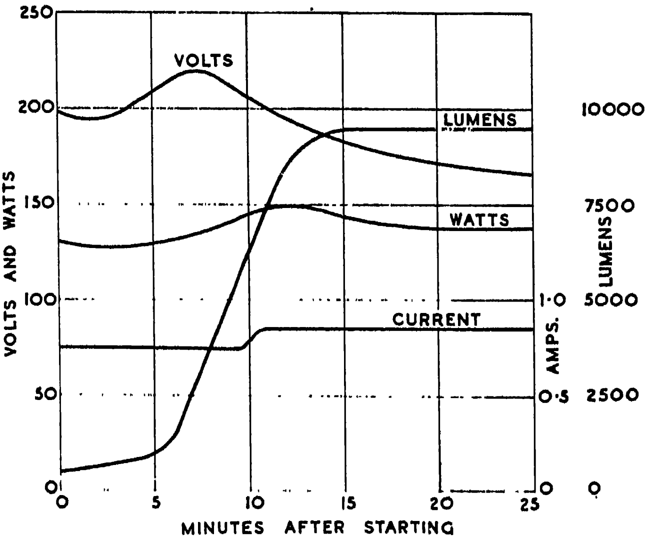


Fig. 19–31

In order to provide this voltage a transformer is used, usually of the auto type. This transformer is provided with a leakage path and functions in a similar manner to those employed for neon lighting. The arrangement of the transformer is shown by Fig. 19–29 and the circuit details by Fig. 19–30.

After the lamp has struck, the restriking voltage at the commencement of each half-cycle is less than the initial striking voltage of 440 volts. Reference to Fig. 19–27 shows that in the case of the 60-watt lamp the restriking voltage is approximately 200 volts.

The starting characteristics of the sodium lamp are quite different from those of the h.p.m.v. lamp. Unlike the latter there is relatively little change in the electrical quantities during the run-up period. The light output, however, is initially very low, but rapidly increases as the sodium becomes vaporized. Some typical characteristics for a 140-watt lamp are shown by Fig. 19–31, from which it will be noted that approximately fifteen minutes are necessary for the lamp to reach its normal operating condition.

Some technical data for sodium lamps are given by Table 19–8.

TABLE 19–8

LAMP WATTS	45	60	85	140
Supply volts . . .	100–250 a.c.	100–250 a.c.	100–250 a.c.	100–250 a.c.
Lamp current, amps. .	0·6	0·6	0·6	0·9
Lumens per watt . .	55·5	65	71·1	71·5
Maximum brightness of arc, candles/cm.^{2} .	9	9	9	9
Overall length, mm. .	250	310	426	550
Arc length, mm. . .	140	170	230	280
Diameter of outer bulb, mm.	50	50	50	65

Fluorescent Lamps

From Fig. 19–3 it is evident that a considerable portion of the radiated energy from mercury-vapour lamps is in the ultra-violet region of the spectrum. By taking advantage of the phenomenon of fluorescence a large proportion of this energy may be employed either for colour correction or improving the efficiency of such lamps. The conversion of electric power to fluorescent light depends on the characteristics of the lamp employed. If the visible spectrum, i.e. 4000 Å to 7000 Å, is divided into ten bands of wavelengths each of 300 Å, the energy radiation of a lamp may be shown in the manner of Fig. 19–32, this being termed a pillargraph. Fig. 19–32 was taken from an 80-watt h.p.m.v. lamp, and it is evident that about as much energy is radiated in the ultra-violet region as within the visible spectrum. Fig. 19–33 shows the radiation characteristics of a l.p.m.v. lamp, and from this the low efficiency of such lamps is apparent and also the very high proportion of ultra-violet radiation in the vicinity of 2537 Å. From Figs. 19–32 and 19–33 it is clear that the possibility of an increase in efficiency by fluorescent means is much greater for low-pressure than for high-pressure lamps. Thus, fluorescent powders are principally employed for colour correction of high-pressure lamps.

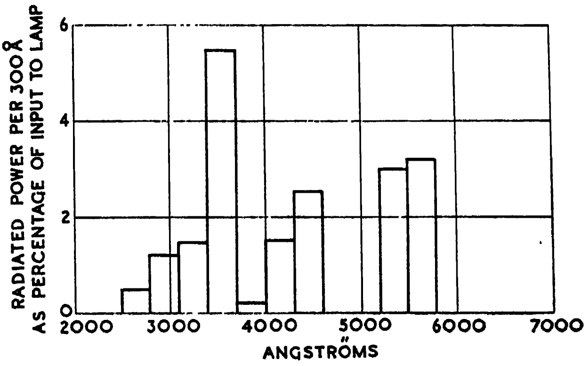


FIG. 19–32

Figs. 19–32 and 19–33 indicate that the ultra-violet power peaks at 3650 Å and 2537 Å for h.p. and l.p. lamps, respectively. Hence fluorescent powders are desirable whose absorption bands also peak at these wavelengths. From Table 7–2 it will be noted that the

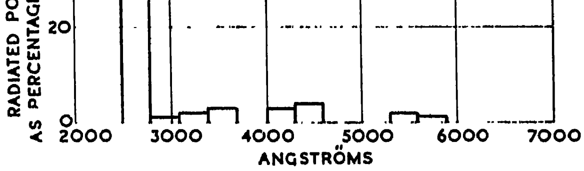


FIG. 19–33

earlier materials of the list are suitable for h.p. lamps, while those lower on the list are more suitable for l.p. lamps. For coating the glass walls of the lamp with the fluorescent powder two methods may be employed. In the first method the wall is pre-coated with a binder and the powder is blown on by a hot-air blast. In the second method the powder is mixed with a binder and coated on the surface. When an organic binder such as glycerine or a nitrocellulose solution is employed, this is subsequently removed by heat. Inorganic binders are dried into a glass-like cement on which the powder is firmly fixed.

An effect of coating a lamp with powder is to absorb about 30 per cent of the light generated within the powder. Thus 70 per cent of the light generated is transmitted and this figure multiplied by the product of the luminous efficiency of the powder and that of generation of ultra-violet radiation of the discharge gives the luminous efficiency of the lamp. As an example, the powder mixture characteristic of Fig. 19–34 may be considered, this characteristic resulting from excitation by radiation of 2537 Å. By summing the percentages of input power in the various wavebands the efficiency is found to be 24 per cent. Multiplying this by the efficiency of generation of ultra-violet power (see Fig. 19–33) and taking 70 per cent of the result to allow for absorption, the overall efficiency is 8·9 per cent.

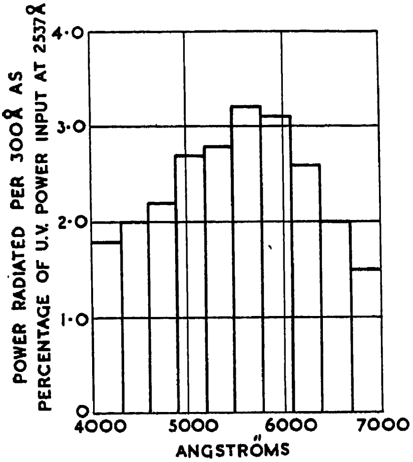


Fig. 19–34

Types of Fluorescent Lamps

The first h.p.m.v. lamp to appear in fluorescent form was a 400-watt type. As the ultra-violet radiation of this lamp peaks at 3650 Å the fluorescent powders employed are ZnS and ZnCdS. These powders are somewhat yellow in colour and therefore tend to absorb some of the blue light from the discharge. This is counteracted by the employment of a small amount of cadmium with the mercury. Because of the high temperature of the discharge tube it is not possible to apply the powder to this. Instead, the powder is applied to the inside of the heat-conserving envelope which, to reduce the temperature still further, is larger than the envelope of the ordinary 400-watt h.p.m.v. lamp. By placing the powder on the inside of the envelope, it is secured from atmospheric contamination. As indicated previously, there is little possibility of improving the efficiency of the h.p.m.v. lamp by means of fluorescence because of the comparatively small amount of ultra-violet radiation present in its spectrum. Due to the absorption of some of the visible radiation by the powder coating, the efficiency of the 400-watt fluorescent lamp is somewhat less than that of the uncoated type, the luminous efficiency of the former being 38 lumens per watt. Some technical data for the 400-watt fluorescent lamp are given by Table 19–9.

TABLE 19–9

Supply volts	200–250 a.c.
Lumens per watt	38
Maximum brightness of arc, candles per cm.^{2}	150
Maximum brightness of bulb, candles per cm.^{2}	50
Length, mm.	335
Arc length, mm.	150
Per cent of input energy radiated	1 at 2600–4000 Å, 9 at 4000–7500 Å
Percentage of red lumens	5

80-Watt and 125-Watt Fluorescent Lamps

These lamps are similar to the h.p.m.v. lamps of the same wattage previously described on page 638 but the pearl-glass bulbs are

TABLE 19–10

Lamp Watts	80	125
Supply volts	200–250 a.c.	200–250 a.c.
Lumens per watt	38	40
Maximum brightness of arc, candles per cm.^{2}	800	800
Maximum brightness of bulb, candles per cm.^{2}	10	15
Length, mm.	178	233
Arc length, mm.	20	30
Per cent of input energy radiated	1 at 2600–4000 Å	12 at 4000–7400 Å
Percentage of red lumens	5	5

replaced by somewhat larger plain bulbs coated with fluorescent powder in a similar manner to the 400-watt lamp described above. Because of the higher arc loading of these lamps, there is more red present, and the degree of colour correction required is not so great as for the 400-watt lamp. Some technical data for both the 80-watt and 125-watt lamps are given by Table 19–10 on page 649.

Low-pressure Fluorescent Lamps

By contrasting the efficiencies of the non-fluorescent and fluorescent lamps of the high-pressure type it is evident that the advantages of the latter lie almost entirely in its superior colour spectrum. This superiority is, of course, due to the employment of the ultra-violet radiation of the lamp, this radiation largely occurring at 3650 Å. Considering the low-pressure lamp, it has already been shown that a very large proportion of the input energy is converted into ultra-violet radiation at 2537 Å. As the efficiency of such lamps is normally low (about 12 lumens per watt) the employment of fluorescence may not only produce colour correction but may also considerably increase the efficiency.

The first low-pressure lamps were of the "neon" type, with a mercury filling. Because of the low operating temperature of such lamps it is possible to apply the fluorescent powder to the inside of the discharge tube, thus obviating the need for an outer envelope. As the ultra-violet radiation of low-pressure lamps lies below 3000 Å the luminescent materials employed with these lamps are those mentioned on page 272, and also given by the latter part of Table 7–2. In the case of neon lamps, the best results are obtained by the employment of $Zn_2SiO_4(Mn)$ and $CaWO_4$, the resulting light being a combination of the red of neon and the colour characteristic of the luminescent materials. Thus, the green fluorescence of Zn_2SiO_4 combined with neon red gives a bright yellow colour, while if $CaWO_4$ replaces Zn_2SiO_4 a pink light results. An important feature of neon-excited fluorescent tubes is the very high lumen maintenance that occurs, the brightness remaining practically constant after three or four thousand hours of use. This is not the case with mercury tubes, where an absorbing film gradually forms over the powder.

On page 263 it was pointed out that for the energy efficiency of luminescence to be high the wavelength of the exciting radiation should not be too small. Thus, if the exciting radiation has a wavelength of 2537 Å and it excites fluorescent light at the optimum visibility of 5550 Å, then assuming $\eta_Q = 1$, the energy efficiency is

2537/5550 = 45·7 per cent. This corresponds to 286 lumens per watt. In the case of neon the exciting radiation has wavelengths of 735 Å and 744 Å and the energy efficiency in this case is 740/5550 = 13·3 per cent, this corresponding to 83 lumens per watt. In the case of h.p.m.v. lamps where excitation occurs at 3650 Å, the theoretical efficiency is 65·8 per cent. The foregoing figures are, of course, theoretical, but serve to show the relative efficiencies that may be expected with different types of fluorescent lamps. In practice the efficiencies are far lower because light is not emitted at the optimum wavelength of 5550 Å and various losses occur, principally heat losses. Table 19–11 shows the order of the efficiencies obtained in practice with various fluorescent materials.

TABLE 19–11

Gas or Vapour	Powder	Colour	Efficiency in Lumens per Watt
Mercury	None	Pale blue	5
	Zinc Orthosilicate	Bright green	50–60
	Zinc Mesodisilicate	Yellow	25
	Calcium Tungstate	Deep blue	15
	Magnesium Tungstate	Light blue	30
Neon	None	Orange-red	15
	Zinc Orthosilicate	Yellow	22
	Calcium Tungstate	Pink	15

The low-pressure fluorescent lamps just described employ cold electrodes and thus require relatively high voltages for their operation. By replacing these electrodes by thermionically emitting electrodes the high voltage necessary to operate the lamp may be reduced to the extent that direct operation from commercial supply voltages is possible. An example of this is, of course, the well-known 5 ft. 80-watt fluorescent lamp. The constructional details of the lamp are shown by Fig. 19–35. It consists of a $1\frac{1}{2}$ in. diameter tube, 5 ft. in length, the ends of which are fitted with ordinary bayonet lamp caps. In each cap are fitted electron-emitting coiled-coil filaments which are arc-heated when the lamp is in normal operation. The gas-filling consists of mercury vapour, and ultra-violet radiation occurs at principally 2537 Å. The fluorescent powders are blended to give a daylight effect, the various constituents having already been described on page 272. A pillargraph of a typical powder mixture employed with an 80-watt fluorescent lamp is shown by Fig. 19–36. By summing the percentages of power radiated over each 300 Å band the efficiency is found to be 13·5 per cent. Multiplying each pillar by the appropriate factor from the luminous visibility curve of Fig. 19–1, a luminous efficiency of 35 lumens per watt is obtained.

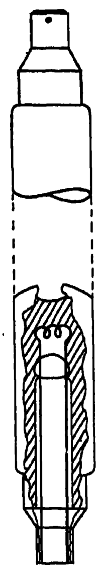


Fig. 19–35

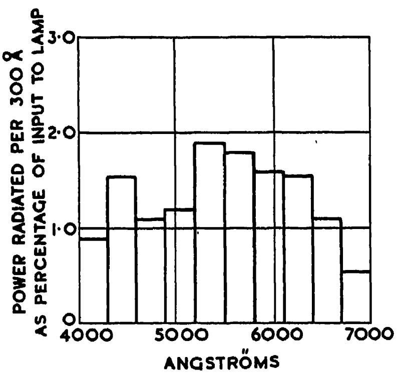


Fig. 19–36

Operating Circuits and Auxiliary Equipment

As with most types of discharge lamps the 80-watt fluorescent lamp is operated in series with a choke. Similar to most low-pressure tubes, and because of its length, a momentary high voltage is necessary to enable the tube to strike. For evident reasons the electrodes must be heated to the point of electronic emission before the application of the striking potential. To pre-heat the electrodes and provide the momentary starting potential, an automatic starting switch is employed, two different types being in general use. These types are known as the "glow" and "thermal" starter switches and are shown in conjunction with the general circuit diagrams of Figs. 19–37 and 19–38. Details of the switches may be appreciated with the assistance of Figs. 19–39 and 19–40. The glow starter

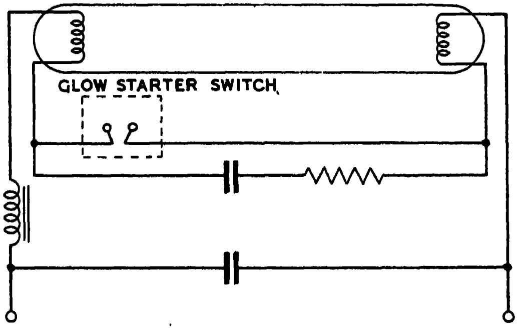


Fig. 19–37

switch, shown by Fig. 19–39, consists of two contacts attached to two bimetal strips, the latter forming the electrodes of a small glow-discharge lamp. On closing the supply switch of Fig. 19–37 the full supply voltage is developed across the starter switch, resulting in a glow discharge. The heat of the glow discharge affects the bimetals, with the result that the contacts close, thus permitting a current to flow to heat the lamp electrodes. Closure of the starter switch contacts, of course, results in extinction of the glow discharge, with the result that the contacts are no longer heated. Thus they reopen and this causes a switching surge, an effect of this being a momentary voltage of about 1000 volts across the lamp. This causes the lamp to strike, whereupon the lamp voltage falls to about 115 volts, which value is too low for a glow discharge to occur in the starter switch. The entire starting process occupies about two seconds.

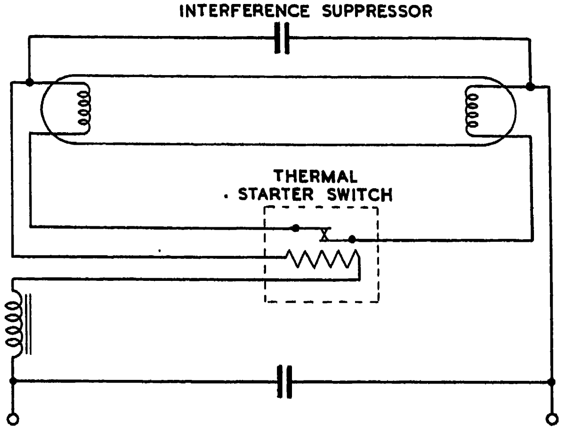


FIG. 19–38

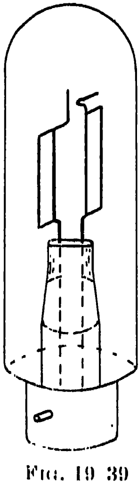


FIG. 19–39

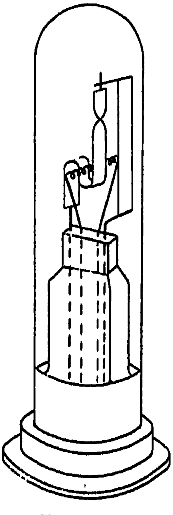


FIG. 19–40

Considering the thermal starter switch of Fig. 19–40, this consists of a pair of contacts, normally closed, supported on bimetal strips.

Close to one of these strips is a small heater. Closure of the supply causes a current to flow through this heater and the lamp electrodes, causing the necessary pre-heating of the latter. The heater of the starter switch rapidly causes the contacts of this to part, whereupon a switching surge causes the lamp to strike. During operation of the lamp the starter switch contacts remain open due to the effect of the heater, the consumption of the latter being about 1 watt.

Lamp Running Characteristics

In contrast with its short starting period of two seconds, the performance of the 80-watt fluorescent lamp is not actually stable for about 15 minutes. At normal air temperatures, however, this is

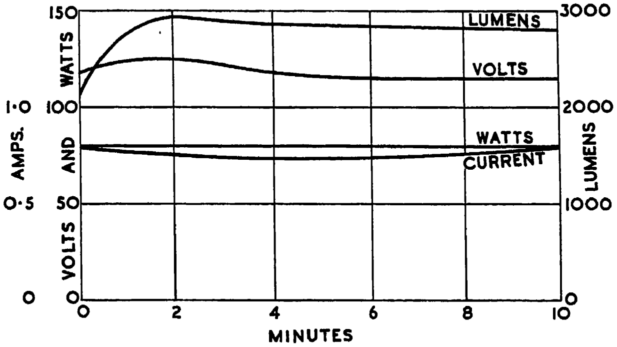


Fig. 19–41

scarcely perceptible as the light output is about 80 per cent of normal at the end of the two-second striking period. Typical lamp characteristics from the end of the striking period onwards are shown by Fig. 19–41. The cause of the somewhat long stabilizing period lies in the low current density of the lamp. Because of this a relatively long period is necessary for the mercury-vapour pressure to rise to the value corresponding to optimum operating conditions. These conditions are obtained when the temperature of the tube walls is from 40° C. to 45° C. It follows that when the ambient temperature falls outside these limits there will be some loss in light output. For example, at 0° C., the initial output is only about 25 per cent of the normal value.

Voltage and Current Waveforms

Figs. 19–42, 19–43, and 19–44, show that the voltage and current waveforms of the 80-watt fluorescent lamp are similar to those of other discharge lamps. The lamp voltage, however, is less constant, as is shown by Fig. 19–42. Also a high-frequency oscillation is in

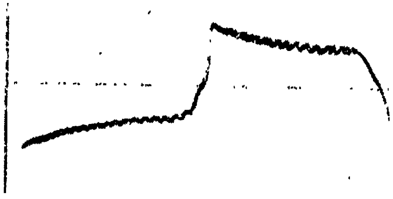


FIG. 19–42

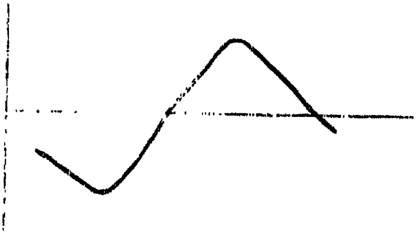


FIG. 19–43

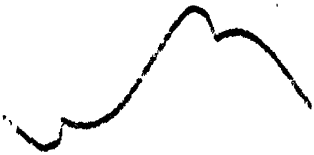


FIG. 19–44

evidence. To suppress this the condenser shown in Fig. 19·38 is connected directly across the lamp terminals.

LAMP SIZES

Although fluorescent lamps can be and are produced having lengths less than 5 ft., the reason for the popularity of this length lies in the relatively high efficiency associated with it. As the cathode loss is independent of length then, evidently, all other factors being equal, the efficiency will increase as the length is increased. Fig. 19–45 shows efficiency plotted as a function of length.

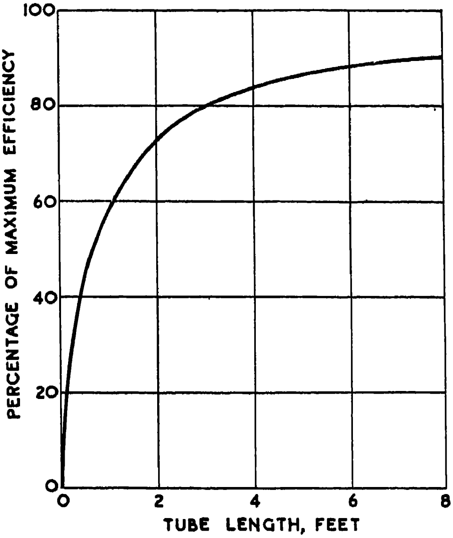


FIG. 19–45

BRIGHTNESS AND COOLNESS

The large surface area of the 5 ft. lamp results in low brightness and freedom from glare, the average brightness being about 4·0 candles per sq. in. A further advantage resulting from the large area is coolness in operation. Thus, in contrast with an incandescent lamp, there is practically no heat radiated to the illuminated surface, the majority of heat generated being dissipated by conduction and convection.

Mathematics of Discharge-lamp Circuits

Because discharge lamps possess a negative resistance characteristic they must be operated in series with some form of impedance. This impedance is either a resistance, a capacitance, or an inductance, and forms of control embodying each of these will now be examined quantitatively.

RESISTIVE CONTROL

In this case control is effected by means of a non-inductive resistance in series with the lamp. At any instant after the lamp has struck the current is given by

$$i = \frac{V_m \sin pt - V_L}{R} \quad . \quad . \quad . \quad (19\text{–}7)$$

where V_L is the lamp voltage (assumed constant), R the value of the resistance, and $p = 2\pi f$, f being the supply frequency. The approximate moments of striking and extinction are, respectively, found from

$$pt_1 = \text{arc sin } V_L/V_m \text{ (striking)}$$

$$(\pi - pt_1) = \text{arc sin } V_L/V_m \text{ (extinction)}$$

From (19–7)

$$i^2 = \frac{V_m^2 \sin^2 pt - 2V_L V_m \sin pt + V_L^2}{R^2}$$

and the r.m.s. value of the current is

$$I = \sqrt{\frac{p}{\pi R^2}\int_{t_1}^{(\pi - pt_1)/p} (V_m^2 \sin^2 pt - 2V_L V_m \sin pt + V_L^2)dt}$$

$$= \frac{1}{R}\sqrt{\frac{1}{2\pi}[(2V_L^2 + V_m^2)(\pi - 2pt_1) + V_m^2 \sin^2 2pt_1 - 8V_L V_m \cos pt_1]}$$

POWER CONSUMPTION

The instantaneous power of the lamp is

$$iV_L = \frac{V_L}{R}(V_m \sin pt - V_L)$$

and the mean power over a cycle is

$$W = \frac{pV_L}{\pi R}\int_{t_1}^{(\pi - pt_1)/p} (V_m \sin pt - V_L)dt$$

$$= \frac{V_L}{\pi R}[2V_m \cos pt_1 - V_L(\pi - 2pt_1)] \qquad (19\text{–}8)$$

POWER FACTOR

Because of the non-sinusoidal character of the current wave, the power factor of a resistive-controlled discharge lamp is less than unity. From (19–7) and (19–8) the power factor is

$$\frac{V_L[2V_m \cos pt_1 - V_L(\pi - 2pt_1)] + \pi R^2 I^2}{V\sqrt{\frac{\pi}{4}[2V_L^2 + V_m^2)(\pi - 2pt_1) + V_m^2 \sin 2pt_1 - 8V_L V_m \cos pt_1]}}$$

which is independent of R and is a function of V_L and V_m only.

The principal objection to resistive control of discharge lamps is the power loss in the resistance. Hence this form of control is only used with small lamps and on direct-current circuits.

CAPACITIVE CONTROL

Control of discharge lamps may be effected by means of a series capacitance. Assuming the circuit to be closed at the moment the supply voltage is passing through a zero value, conduction will not occur until the supply voltage is equal to V_s. While the value of the voltage is less than this we may write

$$V_m \sin pt = V_c + V_L \quad . \quad . \quad . \quad (19\text{–}9)$$

and as the capacity of C, the series capacitance, is large compared with that of the lamp electrodes, (19–9) may be written

$$V_L = V_m \sin pt$$

At the moment the lamp strikes, and during conduction, the current is obtained from differentiation of (19–9). Thus

$$pV_m \cos pt = \frac{d(V_c + V_L)}{dt} = \frac{dV_c}{dt}$$

as V_L is assumed constant. Now $i = CdV_c/dt$ and, hence,

$$i = CpV_m \cos pt \quad . \quad . \quad . \quad (19\text{–}10)$$

From (19–10) it is evident that extinction of the lamp occurs for $pt = \pi/2$ and also that the lamp current is a maximum for $pt = 0$. The maximum value of the current is CpV_m.

PERIOD OF CONDUCTION

At the moment extinction occurs the value of V_c is $(V_m - V_L)$ and conduction does not recommence until

$$(V_m - V_L) - V_m \sin pt = V_L$$

or

$$pt = \text{arc sin}\left(1 - \frac{2V_L}{V_m}\right)$$

Hence conduction persists for a period corresponding to

$$pt_2 - pt_1 = \frac{3\pi}{2} - \text{arc sin}\left(1 - \frac{2V_L}{V_m}\right) \quad . \quad (19\text{–}11)$$

CURRENT, R.M.S. VALUE

From (19–10) the r.m.s. value of the current is

$$I = CpV_m\sqrt{\frac{p}{\pi}\int_{\text{arc sin}\left(1-\frac{2V_L}{V_m}\right)/p}^{3\pi/2p} \cos^2 pt\, dt}$$

$$= CpV_m\sqrt{\frac{1}{\pi}\left(\frac{3\pi}{4} - \frac{1}{4}\sin 2\theta_1 - \frac{1}{2}\theta_1\right)}$$

where $\theta_1 = \text{arc sin}\,(1 - 2V_L/V_m)$.

POWER CONSUMPTION

The instantaneous power of the lamp is

$$CpV_LV_m \cos pt \text{ watts}$$

and the mean power over a cycle is

$$\frac{Cp^2V_LV_m}{\pi}\int_{\text{arc sin}\left(1-\frac{2V_L}{V_m}\right)/p}^{3\pi/2p} \cos pt\,.\,dt$$

$$= \frac{2CpV_L}{\pi}(V_m - V_L) \text{ watts}$$

The power factor is

$$\frac{\dfrac{2CpV_L}{\pi}(V_m - V_L)}{\dfrac{CpV_m^2}{\sqrt{2}}\sqrt{\dfrac{1}{\pi}\left(\dfrac{3\pi}{4} - \dfrac{1}{4}\sin 2\theta_1 - \dfrac{1}{2}\theta_1\right)}}$$

$$= \frac{4V_L(V_m - V_L)}{V_m^2\sqrt{\pi\left(\dfrac{3\pi}{4} - \dfrac{1}{4}\sin 2\theta_1 - \dfrac{1}{2}\theta_1\right)}}$$

As with resistive control the power factor is a function of V_m and V_L only.

Unlike resistive control there are no impedance losses when capacitive control is employed, but a similar disadvantage occurs in that conduction only takes place during a fraction of a cycle. From (19–11) this fraction is

$$\frac{\pi/2 + \theta_1}{\pi}$$

INDUCTIVE CONTROL*

Both the foregoing methods of lamp control suffer from the disadvantage that conduction does not occur over an entire cycle. By employing a series inductance in place of the condenser or resistance an increase in the period of conduction may be effected. This reduces stroboscopic effects and the improvement in the current waveform is usually such that the power factor of the lamp may be raised to little short of unity. During conduction we have

$$L\frac{di}{dt} + iR + V_L = V_m \sin pt \quad . \quad . \quad (19\text{–}12)$$

where R and L are, respectively, the resistance and inductance of the choke. The solution of (19–12) is

$$i = \frac{V_m}{\sqrt{R^2 + p^2L^2}}\left[\sin(pt - \theta) - \varepsilon^{\frac{R}{L}(t_1 - t)} \sin(pt_1 - \theta)\right] + \frac{V_L}{R}\left(\varepsilon^{\frac{R}{L}(t_1 - t)} - 1\right) \quad . \quad . \quad (19\text{–}13)$$

where $\theta = \text{arc tan}\; pL/R$ and t_1 is the time interval between the zero points of the voltage and current waves. Should the current not persist for a full half-cycle, i.e. if the lamp should be struck and extinguished during a time interval of less than half a period, $t_1 = \frac{1}{2\pi f}\text{arc sin}\, V_L/V_m$. If no extinction occurs, i.e. if the current wave is spread over the full half-cycle, then t_1 must be such that

$$-\frac{V_m\left(1 + \varepsilon^{-\frac{R}{2fL}}\right)}{\sqrt{R^2 + p^2L^2}} \sin(pt_1 - \theta) + \frac{V_L}{R}\left(\varepsilon^{-\frac{R}{2fL}} - 1\right) = 0$$

or

$$t_1 = \frac{1}{p}\left[\text{arc sin}\,\frac{V_L\sqrt{R^2 + p^2L^2}}{V_mR}\left(\frac{\varepsilon^{-\frac{R}{2fL}} - 1}{\varepsilon^{-\frac{R}{2fL}} + 1}\right) + \theta\right]$$

As R is invariably negligible, putting $R = 0$ gives

$$i = \frac{1}{pL}\left[V_m(\cos pt_1 - \cos pt) + V_L(pt_1 - pt)\right] \quad . \quad (19\text{–}14)$$

---

* "Choke-Controlled Discharge Lamp Characteristics," F. G. Spreadbury, *The British Engineer*, March, 1940.

and for the case of no extinction

$$t_1 = \frac{1}{p} \text{ arc cos } \frac{\pi V_L}{2V_m} \quad . \quad . \quad . \quad (19\text{–}15)$$

Reference to (19–14) shows the amplitude of the current to be inversely proportional to L while (19–15) shows its time of duration to be independent of L. Assuming the discharge to last for less than half a period, the quantities controlling its time of existence are V_m and V_L. Whether a discharge will persist for less than a half-period may be found by putting $pt = (pt_1 + \pi)$ in either (19–13) or (19–14), pt_1 being equal to arc sin V_L/V_m. If the value of i obtained is negative, the discharge becomes extinguished before the end of the half-cycle; if positive extinction does not occur. It may be noted that the maximum value of the current under the conditions expressed by (19–14) always occurs at the same position relative to the voltage wave, i.e. where $pt =$ arc sin V_L/V_m. Some curves based on the foregoing equations are shown by Fig. 19–46, the data being $V_m = 340$ volts, $V_L = 140$ volts, $L = 16$ henrys, and $f = 50$ cycles per second. The peculiar nature of the choke voltage will be noted, the sharp drop in voltage being due to, and coinciding with, the reversal of lamp voltage. Fig. 19–46 should be compared with the previously given oscillograms of choke voltages. The almost instantaneous change in choke voltage should, of course, be equal to twice the lamp voltage.

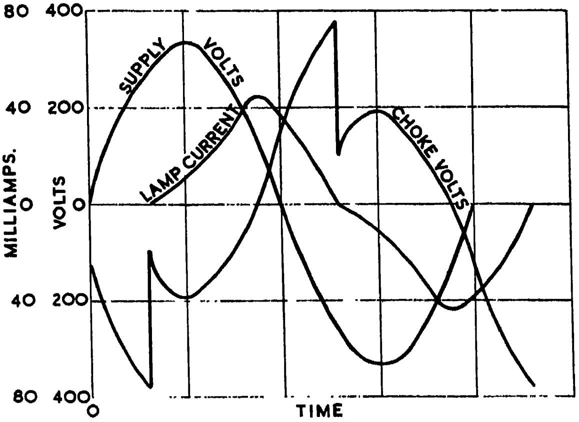


FIG. 19–46

CURRENT WAVEFORM

In most cases the conditions are such that the current in inductively controlled lamps persists for the full half-cycle. This being so, reference to (19–14) shows that the current may be regarded as consisting of two components, one of sinusoidal and the other of triangular waveform. The sinusoidal component is

$$i_s = \frac{V_m}{pL} \cos pt \quad . \quad . \quad . \quad . \quad . \quad . \quad (19\text{–}16)$$

and the triangular

$$i_t = \frac{1}{pL}\left[V_m \cos pt_1 + V_L(pt_1 - pt)\right]$$

The amplitude of the triangular component is $V_m \cos pt_1/pL$ and this maximum occurs when the resultant current through the lamp is zero. Analysis of the triangular component gives

$$i_t = \frac{8V_m \cos pt_1}{\pi^2 pL}\left(\sin pt - \frac{1}{3^2}\sin 3pt + \frac{1}{5^2}\sin 5pt - \text{etc.}\right) \quad (19\text{–}17)$$

The amplitude of the fundamental is 80 per cent of that of the triangular wave and passes through its maximum and zero points at the same moments as the latter. The fundamental, therefore, leads on the supply voltage by $(\pi/2 - pt_1)$ radians. The in-phase component of this fundamental is

$$8V_m \sin pt_1 \cos pt_1/\sqrt{2}\pi^2 pL \text{ amp.}$$

and, as the sinusoidal component given by (19–16) is in quadrature with the supply voltage, it follows that the power taken by the lamp is

$$W = \frac{2V_m^2 \sin 2pt_1}{\pi^2 pL} \text{ watts} \quad . \quad . \quad (19\text{–}18)$$

The 90° leading component of the fundamental is $8V_m \cos^2 pt_1/\sqrt{2}\pi^2 pL$, and this combined with the inductively reactive component $-V_m/\sqrt{2}pL$, gives

$$\frac{V_m}{\sqrt{2}pL}\left(\frac{8\cos^2 pt_1}{\pi^2} - 1\right) \text{ amp.} \quad . \quad . \quad (19\text{–}19)$$

for the resultant reactive component.

CURRENT. R.M.S. VALUE

The r.m.s. value of the lamp current may be expressed by

$$I = \sqrt{I_1^2 + I_2^2 + I_3^2 + \text{etc.}} \quad . \quad . \quad (19\text{–}20)$$

where I_1, I_2, I_3, etc., are the r.m.s. values of the fundamental and its various harmonics. The values of the harmonics may be found from (19–17), but the value of the fundamental in this case must be obtained by taking the vector sum of the fundamental of the triangular component and the inductively reactive component. The phase angle between these two components is $(\pi/2 - pt_1) + \pi/2 = (\pi - pt_1)$ radians. The power factor of the lamp circuit may be obtained by dividing (19–18) by $V_m I/\sqrt{2}$.

From the values given for the inphase component of the fundamental and the total reactive component, the fundamental is given by

$$\sqrt{\left[\frac{V_m}{\sqrt{2}pL}\left(\frac{8\cos^2 pt}{\pi^2} - 1\right)\right]^2 + \left(\frac{8V_m}{\sqrt{2}\pi^2 pL}\sin pt_1 \cos pt_1\right)^2}$$

$$= \frac{V_m}{\sqrt{2}pL}\sqrt{\frac{8^2}{\pi^4}\cos^2 pt_1 - \frac{16}{\pi^2}\cos pt_1 + 1 + \frac{8^2}{\pi^4}\sin^2 pt_1 \cos^2 pt_1}$$

$$= \frac{V_m}{\sqrt{2}pL}\sqrt{1 + \left(\frac{8^2}{\pi^4} - \frac{16}{\pi^2}\right)\cos^2 pt_1}$$

$$= \frac{V_m}{\sqrt{2}pL}\sqrt{1 - (0{\cdot}98)^2 \cos^2 pt_1}$$

$$= \frac{V_m}{\sqrt{2}pL}\sin pt_1 \text{ approximately} \quad . \quad . \quad . \quad . \quad (19\text{–}21)$$

From (19–8) it may be noted that the power taken by a discharge lamp is equal to the lamp voltage multiplied by the mean value of the current. Hence in the present case the power may be obtained by multiplying V_L by the mean value of (19–14). The latter is

$$\frac{p}{\pi pL}\int_{t_1}^{\pi/p + t_1}[V_m(\cos pt_1 - \cos pt) + V_L(pt_1 - pt)]dt$$

$$= \frac{1}{\pi L}\left[V_m\left(t\cos pt_1 - \frac{1}{p}\sin pt\right) + V_L\left(pt_1 t - \frac{p}{2}t^2\right)\right]_{t_1}^{\pi/p + t_1}$$

$$= \frac{2V_m}{\pi pL}\sin pt_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19\text{–}22)$$

Multiplying numerator and denominator by $\cos pt_1$, i.e. by $\pi V_L/2V_m$, the mean value of the current may be expressed as

$$\frac{4V_m^2}{V_L \pi^2 pL}\sin pt_1 \cos pt_1$$

$$= \frac{2V_m^2}{V_L \pi^2 pL}\sin 2pt_1 \quad . \quad . \quad . \quad (19\text{–}23)$$

Multiplying by V_L gives the lamp watts, the resulting expression being identical with that given by (19–18).

POWER FACTOR IMPROVEMENT

The power factor of an inductively controlled discharge-lamp circuit may be improved by removing the reactive component from the line current by means of a condenser. The value of the condenser current is $\sqrt{2}\pi fCV_m$ and the sum of this and (19–19) must equal zero for the power factor to be a maximum. Thus, for maximum power factor,

$$\sqrt{2}\pi fCV_m + \frac{V_m}{\sqrt{2}pL}\left(\frac{8\cos^2 pt_1}{\pi^2} - 1\right) = 0$$

or

$$C = \frac{10^6}{p^2L}\left(1 - \frac{8\cos^2 pt_1}{\pi}\right)\,\mu\text{F}$$

In these circumstances the power factor of the lamp circuit is given by

$$\frac{2\sqrt{2}V_m \sin 2pt_1}{\pi^2 pLI}$$

where the fundamental term in the expression for I is simply the in-phase component of the fundamental of the triangular component.

The various equations previously given for discharge-lamp circuits show that the power factor of such circuits may be considerably less than unity. Thus, with neon and h.p.m.v. lamps, the power factor may be no higher than 0·5, while with sodium lamps figures no higher than 0·3 may be obtained. The low-power factors experienced arise from two causes, the distortion of the current waveform and the lagging or leading effect produced by the condenser or inductance employed for control purposes. In the case of resistive control the low-power factor is, of course, due to the first-mentioned cause. Current distortion occurs for two reasons, firstly, because the current does not persist for a full half-cycle, and, secondly, because a conducting gas behaves as a negative resistance, i.e. an increase in current leads to an increase in gas conductance.

BIBLIOGRAPHY

Electric Discharge Lighting, F. G. Spreadbury (Pitman).
Electric Discharge Lamps, H. Cotton (Chapman & Hall).

CHAPTER XX

MISCELLANEOUS ELECTRONIC DEVICES

The Cyclotron

An important method of investigating the structure of the nucleus of the atom consists of bombarding it with positive ions. This was first accomplished by the employment of high-velocity alpha-particles shot out from radio-active substances. By this means it was found that the lighter elements can be disintegrated. For attacking the nuclei of heavier elements, particles of greater energy than those which are naturally produced are required, and this led to the development of the cyclotron.

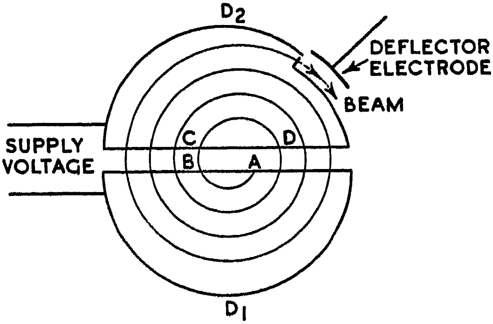


Fig. 20-1

As indicated above, the cyclotron is a device for producing particles of high velocity and energy, the latter in some cases being as high as 4×10^7 electron-volts. Such phenomenal energies are attained with apparatus, the principles of which may be appreciated with the assistance of Fig. 20–1. Referring to this, a high-frequency voltage is applied to two identical electrodes which consist of semicircular shallow metal boxes. These boxes are closed everywhere except at their opposing faces, and because of their structural appearance are often referred to as "dees." Between the dees is a cathode which by virtue of emitted electrons produces ions to be subsequently accelerated. The dees are enclosed within a vacuum chamber, outside of which are the poles of a powerful magnet which produces a field perpendicular to the semicircular faces of the dees. An inlet to the vacuum chamber permits the introduction of small quantities of gas necessary for the production of positive ions.

Consider now an ion which is produced in the gap between the dees. This will be accelerated in the appropriate direction by the electric field in the gap and will, subsequently, enter the interior of, say, dee 1 at A. As the electric field is concentrated mainly between the gap, the ion experiences little or no electric force within the dee. Hence, due to the magnetic field, the path of the ion will be circular and of constant speed until it re-emerges at B. If at emergence the sign of the electric field has reversed, the ion receives a further acceleration and subsequently enters dee 2 at C with a greater velocity than that at which it entered dee 1 at A. Within dee 2 a circular path of constant speed but increased radius will again be pursued, the ion subsequently re-emerging at D. It will be readily appreciated that, providing the alternating voltage is of suitable frequency, the energy of an ion increases at every revolution by an amount up to twice the maximum voltage of the supply. The path of the ion is approximately a planar spiral and, by virtue of many revolutions within the dees, the ion finally approaches their periphery and emerges from this at a suitably arranged outlet with an energy corresponding to millions of volts. Thus, if an ion makes n revolutions and the maximum value of the alternating voltage is V_m volts, the maximum energy with which the ion may emerge is $2neV_m$ electron-volts.

From (3–33) it will be seen that the time of revolution of an ion is independent of its radius and is given by

$$T = \frac{2\pi m}{He} \quad . \quad . \quad . \quad . \quad (20\text{–}1)$$

Hence the frequency of the alternating voltage supply to the dees is

$$f = \frac{1}{T} = \frac{He}{2\pi m} \quad . \quad . \quad . \quad (20\text{–}2)$$

If $H = 10{,}000$ and m is that for the deuteron, then $f = 7{\cdot}75 \times 10^6$ cycles per second. In order to obtain such frequencies it is, of course, necessary to employ a valve oscillator. If R is the radius at which the ions finally emerge, then, from (3–31),

$$R = \frac{mv}{He} \quad . \quad . \quad . \quad . \quad (20\text{–}3)$$

where v is the velocity of emergence. The energy with which an ion emerges is $\frac{1}{2}mv^2$ and from (20–3)

$$\tfrac{1}{2}mv^2 = \frac{1}{2}\frac{e^2}{m} H^2R^2 \quad . \quad . \quad . \quad (20\text{–}4)$$

From this result the importance of using ions of low mass is at once apparent. Also it is evident that the energy may be increased by using larger and larger magnets. The result of this is that cyclotrons employ magnets of enormous mass, that of one of the most recent weighing 2000 tons.

If the ions make too many revolutions they are likely to stray from the desired path, with the result that a proportion of the ionic current may be lost. Hence the voltage on the dees must be high in order, for a given final energy, to keep the number of revolutions low. In practice this voltage may be as much as 10^5 volts. If the maximum voltage is, say, 50,000 volts, and the final energy 10^7 electron-volts, then the ions will make 100 revolutions before they emerge. During their journey, should the ions stray from the median plane, they may strike the dee surfaces before emerging and thus be lost. Hence conditions are so arranged that should an ion stray from the median plane it is automatically brought back. This is effected by causing the magnetic lines of force to curve outwards, as shown by Fig. 20–2. From Fig. 20–2 it will be noted that the field has oppositely directed radial components on opposite sides of the median plane and these will return an ion to the plane should it tend to stray.

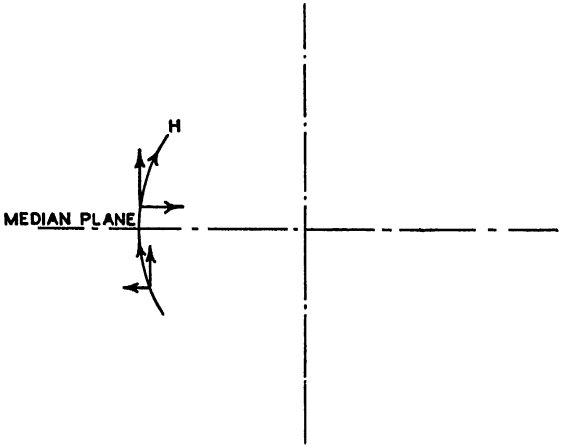


FIG. 20–2

In order to obtain outward curvature of the lines of force as depicted in Fig. 20–2, the field must be decreased in strength in proceeding radially outward. However, from (20–2) it will be noted that the frequency of revolution of the ions varies as the field strength and hence with a radially weakening field the revolution frequency will tend to decrease as the radius becomes larger. Thus, the ions will tend to get out of phase with the electric field as the radius increases and in this manner energy will be lost. It follows that the amount by which the field can be weakened is limited.

Considering the gap between the dees, Fig. 20–3, it is evident that here the electric field forms an electron lens of the two-diaphragm type. Hence, if an electron is approaching the gap in a direction away from the median plane, the focusing action will tend to bring it into this plane.

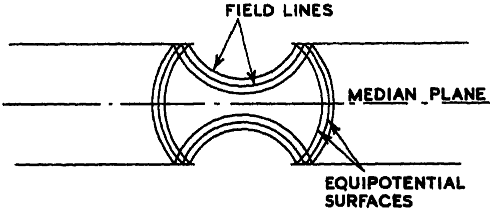


Fig. 20–3

The Betatron

From (20–4) it might appear that higher energy particles would be obtained if electrons were accelerated in the cyclotron rather than ions. However, the final velocity of the ions is given by

$$\tfrac{1}{2}mv^2 = 2neV_m$$

or

$$v = \sqrt{\frac{2E}{m}} \quad . \quad . \quad . \quad (20\text{–}5)$$

where E is the final energy. If E is equal to $10MeV$ amd $m = 2$, then $v = 3{\cdot}1 \times 10^9$ cm. per sec., which is approximately one-tenth of the velocity of light. If the mass of the electron were substituted in (20–4), consideration would have to be taken of its increase in mass with velocity and this, in practice, limits the energy that may be obtained by employing electrons in the cyclotron. Generally, electrons cannot be produced with energies higher than about $0{\cdot}5MeV$.

In order to produce electron energies of the same order as those of ions in the cyclotron, rather different principles are employed and are embodied in an instrument termed the *betatron*. In the betatron a changing magnetic field is employed. If a circular orbit embracing this field is considered, then the e.m.f. round the orbit is proportional to the rate of change of flux within the orbit. Hence an electron on an orbit of radius r experiences an electric force equal to

$$eX = e\frac{d\phi}{dt}\Big/2\pi r \quad . \quad . \quad . \quad (20\text{–}6)$$

where X is the field strength and ϕ the flux. Due to its small mass, the electron tends to travel at enormous speed and during the time the magnetic field is changing may make thousands of revolutions. Because of this it is evident that the electrons must not spiral either inwardly or outwardly during their acceleration. To prevent spiralling a stable orbit, termed the equilibrium orbit, is created at a definite radius, this being achieved by suitable shaping of the magnet pole faces.

The force acting on an electron due to the electric field is

$$\frac{d(mv)}{dt} = eX = \frac{e}{2\pi r}\frac{d\phi}{dt} \quad . \quad . \quad . \quad (20\text{–}7)$$

Now from (20–3) the momentum of an electron travelling in a circular path of radius r is

$$mv = Her$$

Differentiating $$\frac{d(mv)}{dt} = er\frac{dH}{dt}$$

and, from (20–7), $$\frac{e}{2\pi r}\frac{d\phi}{dt} = er\frac{dH}{dt}$$

Integrating, we obtain $$\phi = 2\pi r^2 H$$

or $$\phi/\pi r^2 = 2H$$

From this result it will be noted that to maintain the electron on a fixed orbit of radius r, the average value of H over the orbit must be equal to twice the value of H at the orbit. As stated previously, this is achieved by suitable shaping of the magnet pole faces.

Should electrons tend to stray from the median plane, they are restored in the same manner as in the cyclotron, i.e. by shaping the magnetic field in the fashion shown by Fig. 20–2. A further stabilizing condition is necessary with the betatron. Should electrons attempt to move into orbits of radii different from that of the equilibrium orbit, a restoring force must come into play to return them to the correct orbit. This is achieved by making the magnetic field, H, vary as $1/r^n$, where n is less than unity. The centrifugal force is mv^2/r and the centripetal force Hev, or,

$$\frac{mv^2}{r} = Hev = \frac{Kev}{r^n} \text{ for equilibrium}$$

Putting $$y = \frac{mv^2}{r}$$

$$y_1 = \frac{Kev}{r^n}$$

and differentiating

$$\frac{dy}{dr} = -\frac{mv^2}{r^2}$$

$$\frac{dy_1}{dr} = -\frac{nKev}{r^{n+1}}$$

Hence

$$\frac{dy}{dr}\Big/\frac{dy_1}{dr} = 1/n$$

and, if $n < 1$, the centrifugal force decreases more rapidly than the centripetal force for an increase r and vice versa for a decrease in r. Thus, any change in r brings about a force restoring the electrons to the equilibrium orbit.

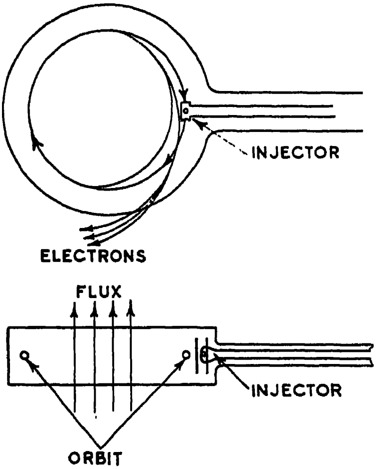


FIG. 20-4

Diagrammatically the operating conditions of the betatron are shown by Fig. 20-4. The magnetic field is alternating in character and is of relatively low frequency, a few hundred cycles per second. The electromagnet acts as an inductance in a parallel resonating circuit, i.e. a bank of condensers is connected in parallel with the betatron magnet winding. The electrons are accelerated in a vacuum chamber placed between the magnet poles, the electrons being injected from an electron gun somewhere near the commencement of the flux cycle. As the flux increases, the electrons are accelerated and must be spiralled out of the equilibrium orbit when the flux is near its maximum value. A period of three-quarters of a cycle then elapses before further electrons are again injected into the vacuum chamber. To spiral the electrons out of their normal orbit towards some desired target, an excess of flux is forced through the orbit near the time of maximum flux.

With the aid of a few assumptions it is possible to obtain some idea of the energy with which electrons emerge from the betatron and also the factors on which this energy depends. Because of the low mass of the electron and the large number of times it revolves in the vacuum chamber, it eventually attains a speed within a few per cent of that of light. If the straight portion of the flux sine curve is employed, $d\phi/dt$ is constant and so also is the force on the electron. Ignoring the change in mass with velocity of the electron, the acceleration is linear and, if v is the final velocity, then the average velocity is $v/2$. The number of revolutions made by an electron is given by

$$\frac{v}{2}T = 2\pi Rn$$

or

$$n = \frac{vT}{4\pi R}$$

The voltage gained per revolution is

$$V = \frac{d\phi}{dt} \times 10^{-8} = \frac{\phi}{T} \times 10^{-8} = \frac{2\pi R^2 H}{T} \times 10^{-8}$$

The final energy is $\quad \dfrac{vT}{4\pi R} \times \dfrac{2\pi R^2 H}{T} \times 10^{-8}$

$$= \frac{v}{2} \times 10^{-8} RH \text{ electron-volts}$$

From this result it will be seen that the final energy of the betatron is proportional to RH, instead of to $(RH)^2$ as with the cyclotron. v, of course, is of the order of the velocity of light.

The Electron Microscope

The limitations of the light-microscope are such that the maximum magnification is about 1000× and the resolving power about 0·2μ (micron). However, even these results are obtained only by the use of ultra-violet light, quartz lenses, and a fluorescent screen for viewing the image. The development of the electron microscope has permitted great advances to be made on the above figures, magnifications up to 200,000× being possible.

The existence of various electron lenses, described in Chapter VI, indicates the possibility of the construction of an electron microscope. In this instrument a beam of electrons is passed through the specimen undergoing examination in a similar manner as a beam of light is passed through a translucent specimen in a light microscope. After the passage of the electron beam, it is focused by electron lenses and a greatly enlarged image of the specimen is produced on a fluorescent screen. Electron microscopes usually possess three lenses equivalent to the glass lenses of the optical system. They may be either magnetic or electrostatic, usually the former, and consist of a condenser lens to focus the electrons on the object, an objective lens to produce a magnified image of the object, and a projecter lens which produces a second magnified image on either a fluorescent screen or a photographic plate.

Reference to Fig. 20–5 gives a comparison between optical and electron miscroscopes, the latter having magnetic lenses. Referring to the former, the condenser lens, L_1, first concentrates the rays from the light source on to the specimen S. The rays, after absorption and refraction by the specimen, then pass through the objective lens, L_2, and a magnified image I is formed. This image is then further magnified by the projector lens L_3 and finally a second further magnified image, I_2, is formed. Turning to the electron microscope, the source of "illumination" is a thermionic emitter. The electrons from this are accelerated by an anode possessing a fine central aperture. Having passed through the latter the electrons come under the influence of the magnetic condenser lens which concentrates them on the object S. The electrons then pass through the specimen, some being absorbed by this according to the density and thickness of its various parts. Those which pass through are then focused by the objective lens L_2 and form a magnified image I. This image forms in turn the object for the projector lens L_3, which finally produces a second further magnified image I_2 on either a fluorescent screen or photographic plate.

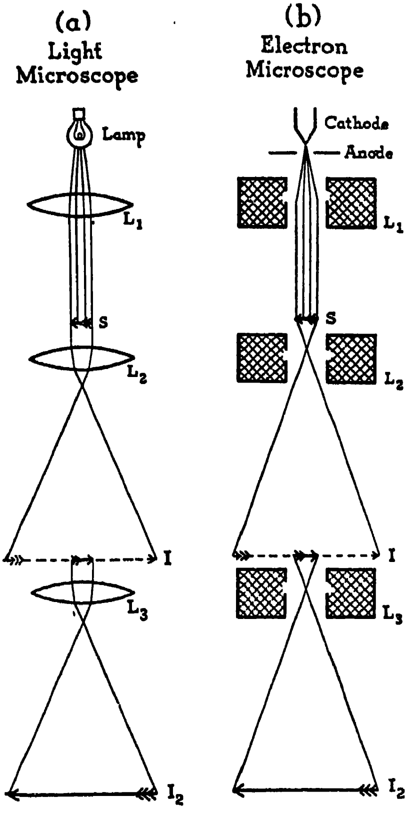


FIG. 20–5

From the formula given for N on page 252, the focal length of a magnetic lens (having the coil form described) is

$$f = \frac{48 \cdot 44 V_0 d}{N^2} \quad . \quad . \quad . \quad (20–8)$$

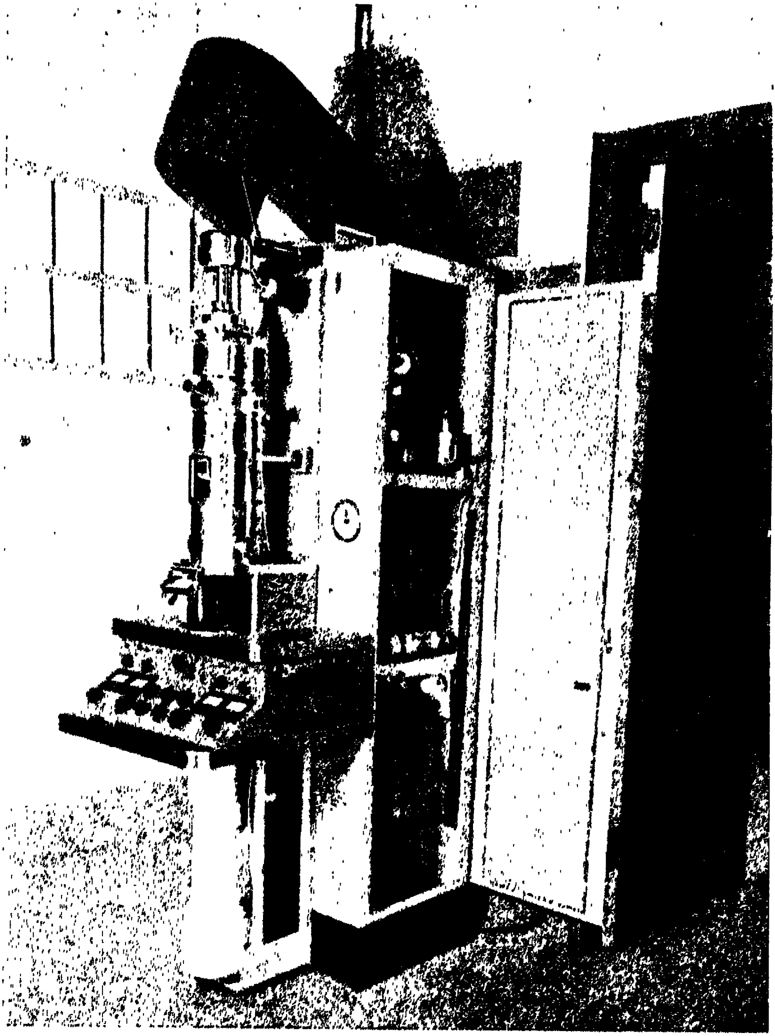


FIG. 20·6

By courtesy of the Radio Corporation of America

For microscopic work it is desirable that f should be as small as possible so that large magnifications may be obtained. In order that a brilliant image shall result, V_0 must be large and is usually between 40 and 80 kilovolts. Hence for f to be small, N must be large and d small. To achieve this the coils are of small internal diameter and are shrouded with iron in a manner similar to that shown by Fig. 6–16. For such coils the right-hand member of (20–8) must be multiplied by a factor, less than unity, whose value depends on the coil geometry. It is evident that the focal length of the lenses, and hence the focusing, of the electron microscope can be varied by adjustment of V_0 and N and, in practice, it is N which is varied. It is, of course, essential that neither V_0 nor N shall be subject to uncontrolled fluctuations and this necessitates closely regulated and smoothed voltage and current supplies. In practice these requirements are met by a high-frequency oscillator operating through filter circuits with negative feed-back regulation.

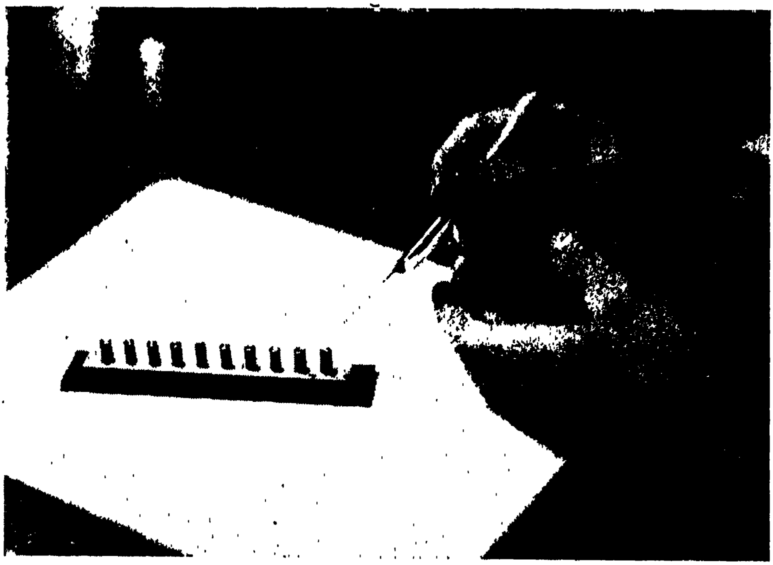


FIG. 20–7

By courtesy of the Radio Corporation of America

The condenser lens of Fig. 20–5 does not need to be of particularly short focal length and hence, although shrouded, does not possess specially-shaped pole pieces as do the objective and projector lenses. Also it is of larger diameter than the latter. For high magnification it has been shown on page 253 that the object must be close to the objective. Hence, in the electron microscope, the object is placed immediately above the magnetic objective. This lens gives a first-image magnification of about 100× and this image is further magnified by the projector lens to give a second-image magnification of about 20,000×. The detail of the final image is such that photographic enlargement may be made up to 200,000×. The focusing of the microscope is effected by adjustment of the objective current, while the magnification is varied by variation of the projector current.

An example of an electron microscope manufactured by the Radio Corporation of America is shown by Fig. 20–6. The electron accelerating chamber and the anode supply are clearly shown at the top, while immediately below is a stainless steel column which forms the body of the microscope and houses the various lenses. Below the control panel is an oil-diffusion pump necessary for evacuating the microscope column to the same order of vacuum as that employed with a thermionic valve, i.e. about 10^{-6} mm. Hg. The meters on the left of the control panel are "coarse" and "fine" ionization gauges. The anode voltage is 50 kV, the various supplies being housed in the cabinet at the rear.

SPECIMEN MOUNTING

In the optical microscope translucent specimens are mounted on glass slides. As substances thicker than about 10^{-6} cm. are opaque to an electron stream, it is evident that a different technique must be employed with the electron microscope. With this instrument a very thin uniform film of collodion or nitro-cellulose is produced showing no structure, this film corresponding to the glass slide in the optical microscope. The film is formed by dropping a small quantity of a 1·5 per cent solution of collodion in amyl acetate on water saturated with amyl acetate. The film spreading over the surface of the water is taken up and dried on a small circular disc of 200-mesh wire gauze, less than $\frac{1}{8}$ in. in diameter. Slight pressure on the gauze causes the film to adhere to it. The coated disc is separated from the rest of the film by means of suitable handling tools, inverted to bring the film side uppermost, and then placed on a miniature pedestal. In this position the water clinging to the surface is removed and the specimen then placed on the film in a drop of fluid, the fluid subsequently evaporating. The whole is then placed at the lower end of a "cartridge" which consists of a tube about 1 in. in length with a bore of about 0·02 in. diameter. A number of pedestals in process of receiving specimen solution are

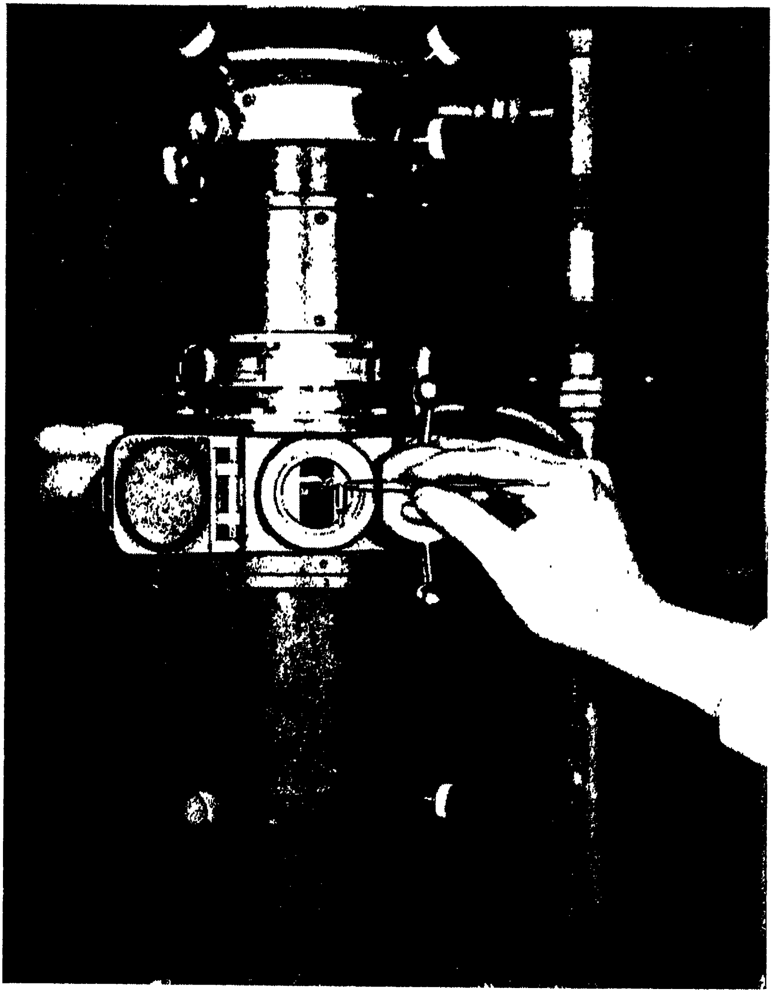


FIG. 20–8

By courtesy of the Radio Corporation of America

shown by Fig. 20–7. The cartridge is placed within the body of the microscope by means of an air-lock (as shown by Fig. 20–8) and the latter then evacuated.

Some typical results obtained with the electron microscope are shown by Figs. 20–9, 20–10, and 20–11. Fig. 20–9 shows soap fibres with a magnification of 45,000×. From this photograph the considerable depth of focus possessed by an electron microscope is apparent. This is an inherent property and, of course, confers on the image a three-dimensional appearance. Fig. 20–10 is the trachea of a honey bee with a magnification of 25,000×. Again the great detail and depth of focus are apparent. Fig. 20–11 is a specimen of polyvinyl chloride with a magnification of 100,000× and demonstrates the remarkable magnifying properties of the electron microscope. It will be noted that the surface is mottled with an evenly spaced succession of spots. The size of these corresponds with that of the unusually large molecules of this material, and it appears that here we have confirmation of the truth of the molecular theory of matter.

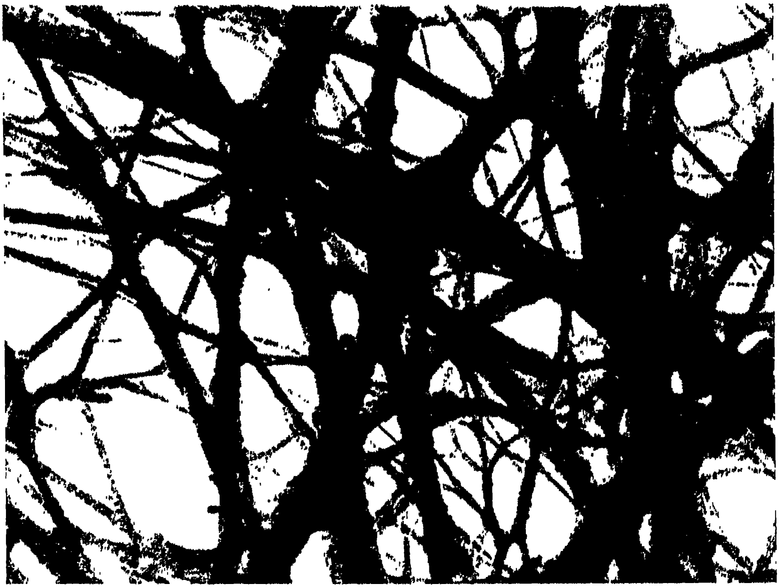


Fig. 20–9

By courtesy of the Radio Corporation of America

For the study of the surfaces of opaque substances a cast is made of the surface, this process being termed the "replica" method. With this method the surface of the metal is first etched and then dipped in a solution of collodion. The collodion forms a film, less than a millionth of an inch in thickness, and on the under side of this film is a reverse replica of the surface contours of the metal. The film is stripped off, mounted in the specimen holder, and examined in the usual manner. As the thickness of the film varies in accordance with the contours of the surface, and the electrons passing through are affected by the degrees of thickness of the film, the resulting image gives a true representation of the metal surface.

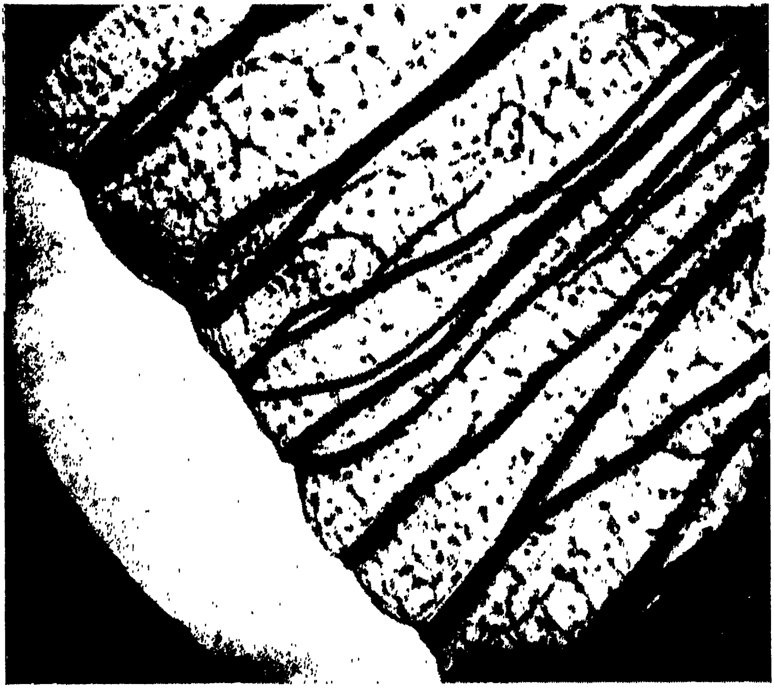


FIG. 20-10

By courtesy of the Radio Corporation of America

The film formed in the above manner is termed a *negative* replica because protrusions above the metal surface appear as thin regions (depressions) in the replica. Where a *positive* replica is desired, a two-stage procedure must be employed. This is carried out in the following manner, using polystyrene as negative material and an extremely thin silica film as the final-replica.

The polystyrene is moulded against the metal at a pressure of

2000–5000 lb. per sq. in. and at a temperature of 160° C. The mould is then allowed to cool when the metal may be removed either by sharply tapping it or, in some cases, by dissolving it. The polystyrene negative is then washed with distilled water and afterwards dried. Following this a thin film of silica is formed on the negative by evaporation. The silica is then removed from all surfaces of the polystyrene, except that which was moulded against the surface of the metal. The silica is now scratched into small squares and the polystyrene then placed in ethyl bromide. This acts as a solvent to the polystyrene and the small squares of silica float off, are captured on specimen screens, and then mounted in the microscope.

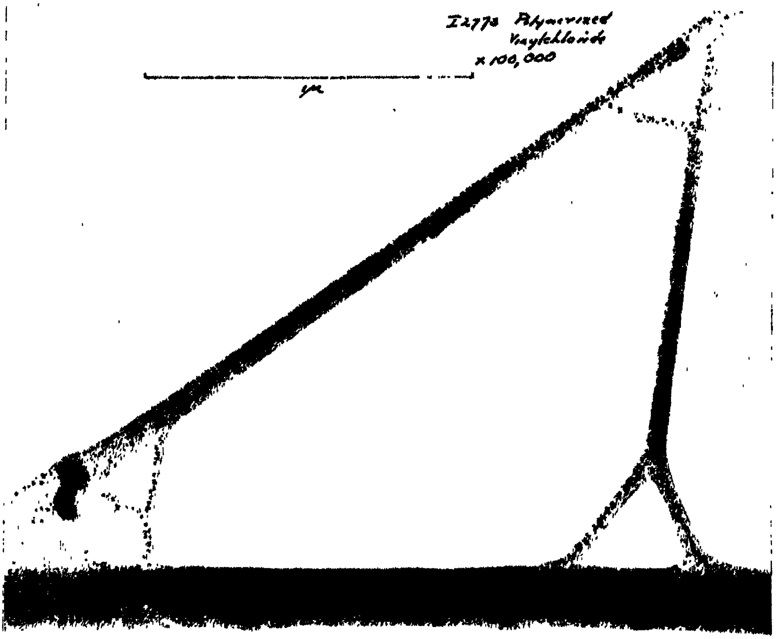


Fig. 20-11

By courtesy of the Radio Corporation of America

The success of the foregoing method of forming positive replicas is based on the enormous mobility of silica over a polystyrene surface. The silica film itself is quite structureless when observed in the microscope and is very transparent to the electron beam.

An example of a polystyrene-silica replica is shown by the photomicrograph of Fig. 20–12. This concerns ordinary steel after being rubbed with fine sandpaper, the magnification being 14,880×. A further example is given by Fig. 20–13, which shows an etched quartz surface. The magnification is 9000×.

An extremely valuable metallurgical application of the electron microscope is the study of electron emitters. Materials under investigation may be used as the microscope cathode and have their images formed on the screen or photographic plate. In this way, emissive properties of various parts of the emitter surface may be studied.

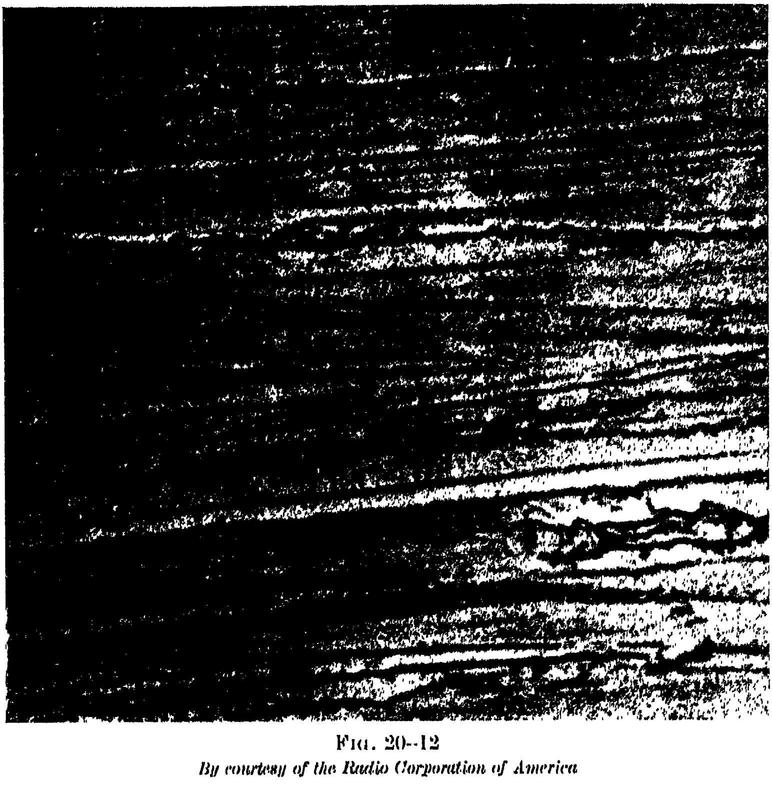


FIG. 20–12

By courtesy of the Radio Corporation of America

LIMITATIONS OF THE ELECTRON MICROSCOPE

Associated with the tremendous advantages of the electron microscope are certain limitations. Thus, finding a given part of the object is of increased difficulty. However, this difficulty has been met either by incorporating an optical microscope to carry out a preliminary search, or by locating the first image before forming the second.

A further difficulty is that, generally, only "dead" specimens can be viewed. This is because the specimen must be located in a high vacuum and is bombarded with high-energy electrons. The effect of the latter is to cause changes in protoplasm and molecular structure. Also, entomologists have recorded shrinkage, evolution of gas, discoloration, and friability of their specimens. A limitation which, of course, has already been mentioned is that the specimen must be extremely thin.

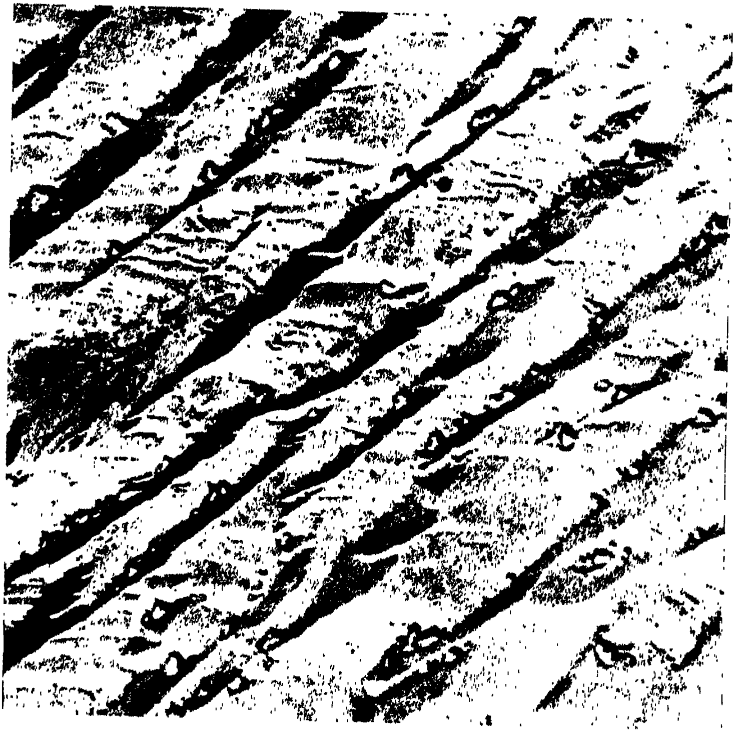


FIG. 20-13

By courtesy of the Radio Corporation of America

The Cathode-ray Tuning Indicator

The cathode-ray tuning indicator, or "electric eye," as it is sometimes termed, offers an indication by virtue of the magnitude of a "shadow" thrown on a small fluorescent screen. The construction of the indicator is shown by Fig. 20–14 and its diagrammatic representation by Fig. 20–15. Referring to Fig. 20–14, the cathode serves as a source of electrons for both the anode of the triode and the target *T*. The latter is conical in shape surmounted by a narrow rim and is coated on its upper side with fluorescent powder. *K* is known as the ray-control electrode and is connected internally to the anode. It is evident that the potential difference between *T* and *K* is equal to the drop on *R*. With no current in *R*, i.e. with the grid biased to cut-off, there is no potential difference between *T* and *K* and no obstacle is offered to a radial flow of electrons from the upper end of the cathode to the target. Hence, under these conditions, no shadow appears on the screen. Now, if the grid is made less negative current will flow in *R* and a potential difference will appear between *T* and *K*. As *K* is negative with respect to *T*, the electrons will be deviated from their radial path in the vicinity of *K* and a shadow will appear on the screen. This shadow is angular in form and is bisected by a projection of *K* on to the target. The smaller the amount of bias on the grid, the greater the potential difference between *T* and *K* and the greater the angle subtended by the shadow becomes. Conditions are usually so arranged that with no grid bias the shadow angle is 90°.

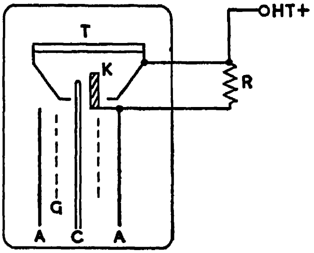


FIG. 20–14

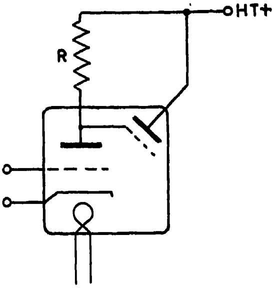


FIG. 20–15

It is evident that the magnitude of the shadow angle is affected by the value of *R* as well as that of the grid-bias voltage. The value of *R* will also affect the value of the grid bias for an absence of shadow. For the case of the Tungsram 6U5 valve, with a target voltage of 250 volts and R equal to 1·0 megohm, the grid bias for no shadow is — 22 volts.

In some tuning indicators, the shadow takes the form of a cross. This is effected by employing four ray-control electrodes spaced at 90° to each other.

The cathode-ray tuning indicator is often used in wireless receivers as an indicator of exact tuning. It has, however, many other applications, a particular case being that of an indicator in instruments such as that described on page 535. If an alternating potential is applied to the grid the shadow angle opens and closes with the frequency of the potential. This results in a sector of reduced shadow density with an angle corresponding to the maximum value of the alternating voltage.

Voltage Stabilization with Discharge Tubes

The fact that the voltage across a glow discharge is practically independent of current enables a glow-discharge tube to be employed as a voltage stabilizer and filter. A simple circuit for improving the regulation of, or smoothing, a d.c. supply is shown by Fig. 20–16. Actually there is a slight increase in voltage with an increase in current through a glow-discharge tube, as shown by Fig. 20–17. If ρ is the slope resistance of the tube and E changes by an amount dE, then the change in tube current is

$$dE/(R + \rho) = dI$$

and
$$dV = \rho dI = \frac{\rho dE}{R + \rho}$$

which, if R is large compared with ρ, becomes

$$dV \simeq \frac{\rho}{R} dE$$

This result at once shows the ability of the circuit of Fig. 20–16 to reduce fluctuations in supply voltages. · Such fluctuations may, of course, exist due to the effect of a load acting on a high impedance supply. Thus, the circuit of Fig. 20–16 may be employed to improve the regulation of a supply and in this connexion it is frequently used.

EFFECT OF LOAD

Let a resistance load R_1 be connected to the discharge-tube terminals and assume that the d.c. supply voltage E contains a sinusoidal component of r.m.s. value e. Then the ripple current due to e is

$$i = \frac{e}{R + \dfrac{\rho R_1}{R_1 + \rho}}$$

also

$$i = \frac{e_1}{\dfrac{\rho R_1}{R_1 + \rho}}$$

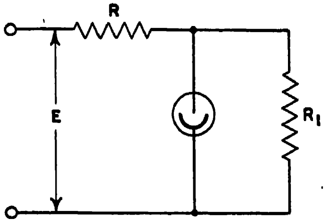


FIG. 20–16

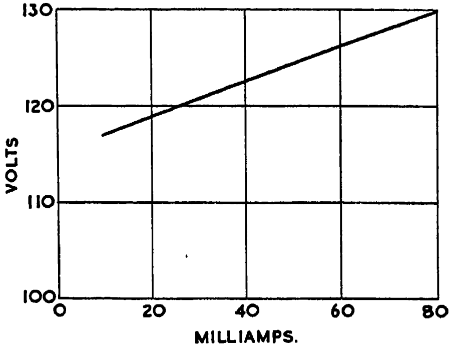


FIG. 20–17

where e_1 is the ripple voltage on R_1.

Hence

$$\frac{e_1}{e} = \frac{\dfrac{\rho R}{R_1 + \rho}}{\dfrac{\rho R_1}{R_1 + \rho} + R} = \frac{\rho}{R} \quad . \quad . \quad . \quad (20\text{–}9)$$

providing ρ is small compared with R and R_1.

REGULATION

Under no-load conditions the current through the lamp is approximately given by

$$\frac{E - V}{R} \quad . \quad . \quad . \quad . \quad (20\text{–}10)$$

where V is the lamp voltage. As the load resistance across the lamp is reduced, the supply current remains roughly constant, while the load current increases. Hence the lamp current falls and reference to Fig. 20–17 indicates that the load voltage also slightly decreases.

Thus the regulation curve is similar to the characteristic of Fig. 20–17. It is evident that the maximum value of the load current depends on the no-load current of the tube and the tube extinction current, being equal to the difference between these two values. A typical regulation characteristic is shown by Fig. 20–18. From (20–9) and (20–10) it is evident that the ripple voltage and maximum load current are inversely proportional to R and hence the degree of smoothing effected by a discharge tube will be smaller the larger the maximum required load current.

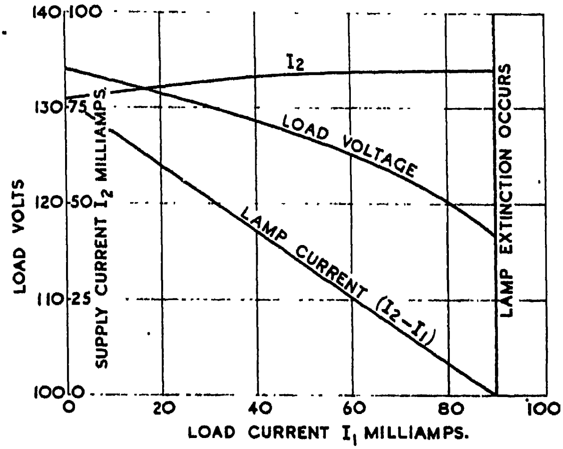


FIG. 20–18

Electronic Voltage Regulators

Several electronic methods exist of regulating the voltage of d.c. generators and alternators possessing considerable advantages over regulators employing relays, vibrating contacts, etc. One obvious advantage is, of course, the rapid action of the electronic regulator due to the negligible inertia of its active parts.

The principle of one method of regulation is indicated by Fig. 20–19. It will be noted that the field of the alternator is supplied from a d.c. source through the triode V_t. Hence variation of a negative bias on this valve will cause variations in the alternator excitation current and, of course, output voltage. Thus the method of regulation consists of suitably varying the bias by means of the alternator voltage. This is effected by means of two other valves, a pentode and a diode, arranged in the manner shown. The triode is biased by the drop on the resistance in the anode circuit of the pentode, V_p, the degree of bias, of course, depending on the value of the anode current, which is, in turn, controlled by the potential of the pentode grid relative to its cathode. The potential of the cathode is maintained at an approximately constant positive value with respect to the negative side of the h.t. supply by means of a neon stabilizer tube V_n. If the alternator is single-phase, then its output voltage is rectified by either a half- or full-wave rectifier, the rectified voltage across R_b being smoothed by the condenser C.

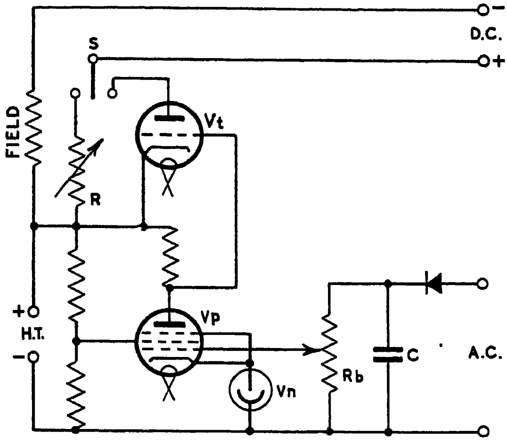


FIG. 20–19

The upper end of R_b is arranged to be positive with respect to the negative side of the h.t. supply, and a portion of the voltage on R_b is applied, as shown, to the control grid of the pentode. The relative magnitudes of the voltages on the cathode and grid of the pentode is such that the potential of the grid is negative with respect to the cathode.

If a three-phase alternator is to be regulated, then three-phase rectification must be employed to produce the necessary grid potential. This, incidentally, tends to maintain the average voltage of the three-phases constant should they be unbalanced. If a d.c. generator is to be controlled, then, obviously, rectification is unnecessary.

Considering now the effect of the regulator on the alternator output voltage, let circumstances be such that the alternator voltage tends to rise. Then the voltage on the pentode grid will increase, thus reducing its negative potential with respect to the cathode.

This increases the pentode anode current and with it the bias on the triode. Hence the triode current is reduced and also the alternator field current which thus reduces the alternator voltage regulation. It will be appreciated that the voltage cannot be held at a strictly constant figure, as some increase is essential in order to increase the bias on V_t. Nevertheless, a considerable improvement in regulation may be effected with the arrangement shown.

It is apparent that if the valve heater supplies, etc., are derived from the regulated a.c. source, then the alternator must first be independently excited in order that the regulator may come into operation. This is effected by the switch and variable resistance shunting V_t. Initially S closes the resistance circuit, R being adjusted so that the alternator voltage builds up to its normal value. At this point the switch is changed over and V_t substituted for R.

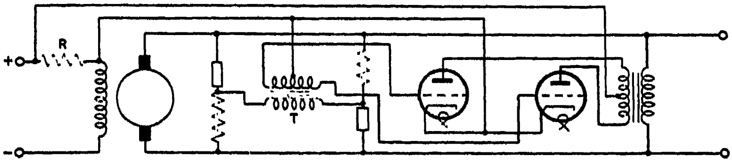


FIG. 20–20

S should be so constructed that the valve is brought in before the resistance circuit is broken.

A further method of alternator regulation, employing thyratrons, is shown by Fig. 20–20. It will be noted that a bridge circuit is employed, this being similar in character to that of Fig. 20–21, i.e. it consists of two arms having linear resistances and two having non-linear resistances. Such bridges balance for only one value of the applied voltage and conditions in the present case are arranged to be such that this voltage is that at which regulation is desired. Assuming the bridge to be balanced there will be no e.m.f. across the primary of the transformer T and, consequently, the cathode-grid potential difference of the thyratrons will be zero. Under these conditions the valves are conducting for about half the permeable half-cycle and a certain current is delivered by the thyratron rectifier to the alternator field circuit via R. Assuming now that a rise in alternator voltage tends to take place, the bridge will unbalance and a cathode-grid potential difference will occur. The connexions of T are such that the grid potentials now become negative during the conducting half-cycle, this decreasing the rectifier output and thus reducing the alternator field current. Should the alternator voltage tend to fall, the grids become positive, increasing the conduction period and hence the excitation current. It will be appreciated that the alternator regulation may be considerably improved by this device.

A further method of regulation, employing both vacuum tubes and thyratrons is shown by Fig. 20–21. A bridge is again employed similar in character of that of the previous regulator. In this instance, however, it is operated from d.c. derived from a single-phase full-wave vacuum-type rectifier connected between two lines. The output from this rectifier is smoothed by L and C, and then applied to the bridge, which is arranged to be balanced when the alternator voltage is at its normal regulated value. R is, of course, for the initial setting of the regulator.

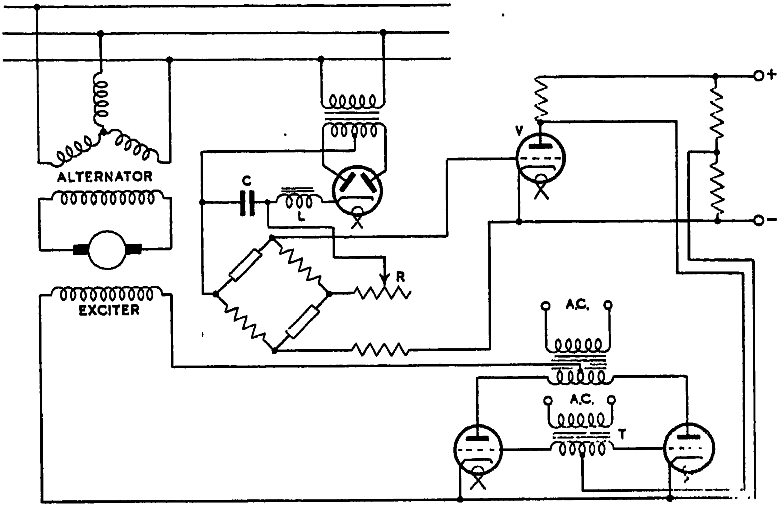


Fig. 20–21

The out-of-balance voltage of the bridge is applied between grid and cathode of V and is amplified by this valve, the amplified output voltage being combined with a constant voltage from the transformer T. We thus have bias-shift control of the thyratron output as described on page 495. Assuming the alternator voltage to rise, the connexions of the regulator are such that the negative bias is increased, thus reducing the thyratron output and the input to the exciter field. Should the alternator voltage fall, the reverse effect occurs.

It is found that with the last method of regulation less than 1 per cent change in alternator voltage may produce 100 per cent change in exciter field current. Hence it is clear that a very close degree of regulation may be achieved.

Electronic Control of Temperature

An electronic method of temperature regulation, employing hard valves, is indicated by Fig. 20–22. On the extreme left is a resistance type thermometer, two arms of which are formed by centre-tapping the secondary winding of a mains-operated transformer. One of the remaining two arms consists of a variable resistance, having a negligible resistance temperature coefficient, and is used for balancing the bridge at any required temperature. The other arm, of course, constitutes the thermometer. The out-of-balance voltage of the bridge is applied between grid and cathode of the valve V_1, where it undergoes amplification, the amplified voltage then being applied to the valve V_2. This valve acts as a half-wave rectifier, being supplied from the transformer T_2. The d.c. output voltage is developed across R_3, which is smoothed by C_1, R_4, and C_2, and applied between grid and cathode of V_3. In the anode circuit of V_3 is a d.c. relay which controls the power supply to whatever is undergoing temperature control.

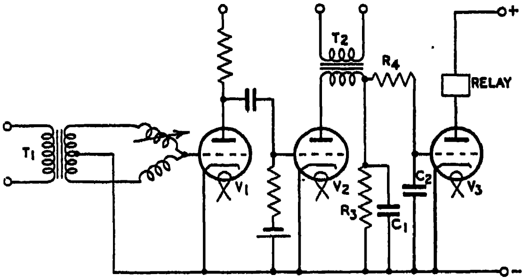


FIG. 20–22

Let it be assumed that the bridge is initially balanced and that the temperature then falls. This will unbalance the bridge and an alternating e.m.f. will appear between grid and cathode of V_2. Now if the secondary of T_2 is suitably connected, the voltages on the anode and grid of V_2 will be in phase opposition and a falling temperature will decrease the anode current of this valve, for when the anode becomes positive the grid will become more negative. The consequent reduction in voltage on R_3 reduces the bias of V_3, with a consequent increase in the anode and relay current of this valve. Hence the relay may function and increase the heating power, thus restoring the temperature to the balanced value.

Should the temperature rise above the balanced value, a phase reversal of the out-of-balance voltage from the bridge will occur. Thus the phase relationship of the anode and grid potential of V_2 will be reversed and the anode current and voltage across R_3 will increase. This will increase the bias on V_3, reducing the anode and relay current of this valve. Hence the relay may function and decrease the heating power, thus restoring the temperature to the balanced value.

Consideration of the circuit details of Fig. 20–22 will reveal the necessity of R_3, for in its absence the rectifier would work into an open circuit and once C_1 and C_2 become fully charged. the grid potential of V_2 would be without effect.

BIBLIOGRAPHY

"The Cyclotron and Betatron," J. D. Cockcroft, *J. Sci. Inst.*, Nov., 1944, p. 189.

"Historical Development of the Betatron," D. W. Kerst, *Nature*, 26th Jan., 1946, p. 90.

"Production of Particle Energies beyond 200 MeV, L. T. Schiff, *Review Sci. Inst.*, Jan., 1946, p. 6.

"Biological Applications of the Electron Microscope," G. E. Donovan, *Nature*, 16th Sept., 1944, p. 356.

"Voltage Stabilizer," G. N. Patchett, *Elec. Review*, 28th April, 1944, p. 602.

"A Wide Range Sensitive Thermoregulator," G. E. Coates, *J. Sci. Inst.*, May, 1944. p. 86.

"Bridge Voltage Regulators," C. Morton, *J. Sci. Inst.*, Jan., 1944, p. 15.

SUBJECTS INDEX

ABBE'S Sine Law, 242
Action, 33
Activation, 267, 284, 285
Activators, 267
After-glow, 260, 271
Ageing, 406
Alcomax, 519
Alpha-particle, 35, 520
Aluminium oxide, 286
Amplification, 338
—— factor, 313, 315, 316
Amplifier, 520
Amplitude distortion, 350
Anode—
  bend detection, 324
  characteristics, 352
  excitation, 424
  impedance, 316
  loading, 472, 479
  materials, 288
  rectifier, 428
  seals, 430
  sleeves, 429
  tapping point, 389
  temperature, 301
Antinodes, 257
Atom, 2
Atomic number, 35
—— radius, 2, 35, 37
—— susceptibility, 143
Atomistic concepts, 1
Auto-electronic emission, 92
Average speed, 9
Avogadro's Law, 15

BACK-FIRE, 414
Back-wall cell, 611
Baird photo-cell, 605
Balmer series, 38
Barium, 285
Barrier layer, 402
Beam trap, 551
Beryllium oxide, 286
Binder, 285
Black body, 24
—— —— radiation, 26
—— —— radiator, 623
Blacking out, 560, 571
Blocking layer, 402
—— potential, 489
Bohr's postulates, 36
Boltzmann's constant, 153, 165

Bound electrons, 159, 204, 210
Boyle's Law, 4, 12
Brackett series, 39
Breakdown, 291
Bridge network, 526, 529

CADMIUM SULPHIDE, 267
—— tungstate, 267
Caesium, 543
Calcium molybdate, 267
—— sulphide, 261, 267
—— tungstate, 267
Carnot cycle, 155
Cartestian co-ordinates, 32
Cathode—
  antimony-caesium, 610
  arc-heating of, 295
  disintegration, 295, 299
  efficiency, 283
  emission, 299
  evaporation, 295
  fall, 87, 291
  ——, abnormal, 91, 626
  heat-shielded, 300
  injection, 336
  spot, 92, 299, 425
  temperature, 283, 284
Centre-tapped winding, 280
Choke, 342
——, lamp, 637
Class B valve, 334
—— —— transformer, 377
Clean-up, 90, 220, 625
——, valve, 297
Coefficient of emissivity, 288
Cold light source, 620
Collision, 3
—— frequency, 16
Colpitts circuit, 392
Combination frequencies, 325
Composite deflexion, 559
—— dynamic characteristics, 373
—— static characteristics, 371
Compound, 2
Compton scattering, 203
Condensation chamber, 429
Condenser charging time, 561, 564
Contact breaker, 502
—— potential, 44, 591
Control ratio, 302
Corona, 76
Coronal extension, 73

Coulomb force, 35
Counter electrode, 404, 610
Coupling condenser, 339
——, loose, 379
——, tight, 379
Creep, 406
Critical potential, 44
Crookes dark space, 86, 89, 99
—— tube, 197
Cross-firing, 411
Cuprous oxide, 196, 403, 610
Curie point, 146
Curie-Weiss Law, 146
Current number, 38
Cut-off, 374, 378
Cycloid, 128
Cylindrical co-ordinates, 130

DARK CURRENT, 617
—— space, 85
De-activation, 286
Dees, 665
Deflexion factor, 554, 556
Defocusing, 546
De-ionization, 90, 513
—— time, 302, 309
Depth of modulation, 323
Deuteron, 666
Dielectric heating, 137
—— polarization, 135
Diffraction pattern, 213
Dipole, 134, 139, 146
Discharge—
  brush, 72
  corona, 72
  glow, 298
  self-maintained, 64
  separately-maintained, 62
Disintegrating voltage, 299
D-lines, 50
Doublet, 50
Drift velocity, 81
Driver valve, 374
Dynamic characteristic, 350
Dynamometer wattmeter 537
Dynatron, 331, 395

EFFICIENCY—
  quantum, 263
  luminous, 263
Einstein's equation, 190
Electric eye, 682
Electrode, activated, 632
Electrodes, 90
Electro-luminescence, 623
Electrolytic trough, 243
Electrometer, 58
—— capacity, 59
Electro-positive metal, 192
Electron—
  assignment, 51
  avalanche, 82
  beam diameter, 255
  coupling, 337
  definition, 2
  density, 166, 282
  diffusion, 82
  energy, 168
  inertia, 103, 111
  multiplication, 605
  planetary, 34
  radius, 112, 122
  recoil, 205
  secondary, 604
  shells, 3, 51
  spin, 165
  temperature, 96
  trajectory, 244
  transit time, 556
  velocity, 101
  -volt, 46
Electroscope, 59
Element, 2
Ellipse, 579
Emission cut off, 591
—— efficiency, 283
——, total, 284
Energy distribution formulae, 25
—— level, 635
—— —— diagram, 39, 162
Equipotential cathode, 279
Equivalent diode, 314
Evacuation, 285
Excitation, 42
—— potential, 42
Excited, 42
Exposure time, X-ray, 223

FARADAY dark space, 89
Feed, 392
Feedback—
  compound, 383
  current, 381
  voltage, 382
Field emission, 92
—— theory, 1
Figure of merit, 58
Filament vaporization, 298
—— voltage, 300
Fine stucture, 50
Fire, 291
First detector, 335
Floating potential, 94

Fluorescence, 197, 259
Fluorescent lamps, 271
—— powder, 648,651
—— radiation, 207
Fluoroscopy, 266
Flyback, 560
—— velocity, 560, 565
Focal length, 232
Forward resistance, 400
Free electrons, 103, 149, 157, 159, 402
Frequency—
  comparision, 582
  distortion, 362
  fixing, 514
Front-wall cell, 611

GAIN, 339, 379
Galvanometer, 58
Gap—
  parallel wires, 75
  point and plane, 74
  wire and cylinder, 73
Gas—
  amplification, 604
  breakdown, 65, 68, 70
  conductivity, 57
  density, effect of, 75
  -discharge triode, 301
  -focussed tube, 256
  ideal, 4
  inert,625
  perfect, 3
  permanent,4
  pressure, thyratron, 310
  ——, valve, 297
  real, 4
Generalized co-ordinates, 31
Gold-leaf electroscope, 57
Gradient, 73
Grid—
  bias, 392
  ——, optimum, 362
  current, thyratron, 307
  detection, 328
  -excitation transformer, 504
  leak, 340
  suppressor, 333
  thyratron, 308
  vacuum valve, 312
Ground state, 42

HARDENING, 625
Harmonic, 537
—— distortion, 350, 379
Hartley circuit, 391
Heater, 287
Heat loss of emission, 176
Heated-shielded cathode, 287
Helium, 40
Heptode, 337
Holding-arc, 424
Hum, 280
Hydrogen spectrum, 38

IGNITION—
  angle, 490, 498
  electrode, 424
  ignitron, 433
  potential, 294
Image—
  force, 184
  rotation, 249
Impulse ratio, 76, 78
Impurities, 267
Impurity levels, 196, 614
Incremental permeability, 347
Inner photo-electric effect, 616
Inter-electrode capacity, 319
Interphase rating, 477, 486
—— —— reactor, 474
Inversion, 506
Ion—
  definition, 42
  mobility of, 61
  velocity of, 60
Ionization by collision, 62
—— potential, 42
Ionized, 42
Ionizing agents, 59
Iron selenide, 611

KILLERS, 269
Kinetic energy, 32
K-series, 210

LANGMUIR'S equation, 276, 290
Larmor precession, 53
Latice stucture, 27, 152, 157
Leakage inductance, 345
Lecher wires, 397
Length of scan, 564
Liberating potential, 489
Light quantum, 189
Line-focus tube, 221
Load line, 352, 602
L-series, 210
Lumen, 620
Luminescent centres, 265
Luminous discharge tubes, 84
—— efficiency, 266, 625
Lyman series, 38

MAGNET, cyclotron, 666

Magnetic susceptibility, 146
Magnetron, 134, 397
Mass-absorption coefficient, 202
Mass-point, 32
Matching transformer, 363
Maximum undistorted power, 360
Maxwell's Law of Velocity, 7, 24, 76
Mean free path, 3
Mercury vapour—
  critical potentials, 45
  influence of temperature on, 296
  lamp, 271
  temperature/pressure characteristic, 428
  use with high inverse voltages, 298
Metastable state, 260
Microscopic concepts, 1
Miller indices, 213
Minimum sparking potential, 69
Mixer, 335
Molecule, 2
Molecular chaos, 4
—— speeds, 11
Momentum co-ordinates, 32
Monochromatic radiation, 42
Most probable speed, 9
Multiplets, 50
Mumetal, 347
Mutual characteristics, 313.
—— conductance, 287, 317

Negative—
  corpuscles, 100
  glow, 87
  grid current, 303
  resistance, 385
  —— oscillator, 395
  space charge, 273, 283
Neon lamp, 271, 560
Nodes, 257
Non-linear distortion, 350
Nucleus, 2, 34, 158
——, motion of, 39

Octode, 336
Ohm's Law, 150
Optical characteristics, 634
Origin distortion, 545
Oscillation amplitude, 394
Oscillator—
  efficiency, 389
  energy, 29
  stability, 394
Oscillogram, 631, 641, 644, 655
Oscillograph plate impedance, 544
Overlap, 424, 444
Oxidation, 403
Ozone, 72

Parallel feed, 347
Paschen's Law, 66, 295, 412
—— series, 39
Peltier effect, 181
Penetration factor, 314
Pentode, 332
Permalloy, 347
Permanence of luminescence, 266
Persistence of vision, 563
Phase—
  integral, 34
  path, 33
  plane, 33
  shift transformer, 497
  shifter, 492, 494
  space, 163
  splitting, 580
Phosphorescence, 259
——, persistence of, 260
Phosphorogens, 267
Photo-electric efficiency, 590
Photo-electrons, 77
Photo-excitation, 43, 66
Photometric characteristics, 633
Photon, 205
Pillargraph, 646
Planck's constant, 33, 189
—— formula, 29
Plasma, 83, 87, 94, 291, 298, 300, 304
—— characteristics, 90
PM6 valve, 528
Poisson's equation, 274, 275
—— Law, 89
Polar co-ordinates, 32
Position co-ordinates, 32
Positive—
  column, 87, 90, 627
  grid drive, 374
  ion current, 303
  —— bombardment, 283
  —— density, 282
Post-accelerating electrode, 554
Potassium, 593
Potential barrier, 159, 184
—— energy, 32
—— minimum, 273
Power sensitivity, 366
—— supply, 582, 584
Precession, 141
Primary current, 612
—— electrons, 191
—— X-rays, 198
Probability curve, 8
—— of collision, 18
Probe, 414

Pump, 432

Q-POINT, 352
Quanta, 25
Quantum efficiency, 590
Quantum number—
  azimuthal, 47
  definition, 36
  magnetic, 49
  radial, 47
  spin, 51
Quantum yield, 590
Quartz, 591, 639
—— lenses, 671
Quenching, 268
Quiescent current, 351

RADIAL current, 90
Radiant efficiency, 662
Random currents, 90, 95
—— velocity, 150
Raster, 548
Reaction, 385
Recombination, 60, 310, 630
Rectangular co-ordinates, 32
Rectification, 323
——, grid, 325
—— ratio, 401
Rectifier—
  capacities, 399, 423
  cell, 610
  compounding of, 506
  cooling, 430
  excitation, 432
  formation, 409
  glass bulb, 424
  ignition, 432
  resistance, 400
  reverse losses, 438
  steel bulb, 422
  —— tank, 421, 429
  temperature, 407
  triodes as, 322
  volt drop, 451
  voltage regulation, 480
  —— ripple, 457, 480
Refractive index, 136
Regeneration, 510
Relative velocity, 17
—— visibility factor, 620
Relativity, 117
Relaxation oscillator, 395
Replica, 677
Resolving power, 671
Retarding potential, 591
Reverse resistance, 400
Ripple, 685
Rubidium, 543

Running levels, 195, 616
—— state, 195
Rydberg constant, 39, 41, 219

SANDWICH cell, 610
Saturation current, 153, 184, 187, 273
Saw-tooth oscillator, 560
Scattered radiation, 204
—— X-rays, 198
Scattering angle, 206
Schottky effect, 592
Second harmonic, 324
—— —— distortion, 354, 369
Secondary electrons, 191
—— emission, 65, 264, 330
—— X-rays, 198
Selenium, 408
Self-bias, 392
Self-capacitance, 345
Sensitivity, 553, 556
——, dynamic, 558
Sensitization, 593
Shadow angle, 535, 682
Sheath, 98, 291, 305
—— thickness, 306
Signal amplification, 585
Signal-to-noise ratio, 609
Sinusoidal oscillator, 395
Slope resistance, 316
Sodium, 593, 642
Sound-film reproduction, 608
Space charge, 76, 85, 273, 294
—— —— density, 84
—— —— equation, 275
Specific heat, 160
Spectral sensitivity, 590
Speech coil, 363
Sphere of influence, 3, 16
Spherical aberration, 231, 252
Sputtering, 90, 610, 625
Stage gain, 338
Standing figure, 563
Starter switch, lamp, 652
Stationary, 36
Statistical equilibrium, 4
Stefan-Boltzmann Law, 31, 288
Stoke's Law, 109
Streamer, 79, 81
Strike, 291
Striking potential, 294
Strontium, 285
Superheterodyne, 334
Survival equation, 19
Switching surge, 653
Symmetrical deflexion, 549
Synchronization, 563, 573
——, tight, 579

TARGET, 220, 682
Tellurium, 616
Temperature—
  lamp, 635, 643, 654
  radiation, 259
  radiator, efficiency of, 620
  rise, valve, 297
Tertiary winding, 457
Tetrode, 322, 329
Thallous oxysulphide, 616
Thermionic emission, 91, 168
Thermodynamics, 25
Third-harmonic distortion, 356
Third-point spark gap, 78
Thoria, 264, 284
Thorium oxide, 284
Three-halves power law, 274, 313, 316
Threshold frequency, 589
—— voltage, 264
—— wavelength, 264
Thyratron, voltage rating, 311
—— voltmeter, 525
Tilted electroscope, 58
Time base—
  condensers, 584
  efficiency, 564
  frequency limit, 567
  Puckle's 569
Townsend coefficient, 80
—— theory, 80
Transformer—
  auto, 645
  -coupled amplifier, 343
  reactance, 450
Transition current, 479
Trapezium distortion, 548
Triggering circuit, 568
Triode-pentode, 335
Triplet, 50
Tungar rectifier, 295
Tungsten, 283
——, thoriated, 283

VACUUM, 281
——, rectifier, 432
Valve—
  argon-filled, 412
  B.D.12, 300
  efficiency of, 365
  heating time, 301
  rating, 299
  6U5, 536, 682
Vapour pressure, 106
Velocity-point, 4
Velocity-space, 4
Vibrator-converter, 514
Virtual cathode, 553
Voltage, extinction, 630
Voltmeter, moving coil, 519
—— zero error, 532
Volume control, 334
Volume-element, 21

WAVE number, 39
Wehnelt cylinder, 543, 547
Wein's Displacement Law, 30
—— Law, 27
Work function, 170, 283, 590
—— voltage, 573

NAMES INDEX

AITKEN, 104, 105
Ardenne, von, 546
Avogadro, A., 15, 157

BALMER, J. J., 38
Barkla, C. G., 207
Bloch, E., 56
Blumlein, 566
Bohr, N., 36
Boltzmann, L., 31
Bolye, R., 4
Brackett, P., 39
Bragg, W. H., 218, 226
Brockelsby, C. F., 541

CAGE, J. M., 434
Cance, J. C. R., 309
Child, C. D., 274
Clarke, G. L., 222
Clerk-Maxwell, J., 7, 170
Coates, G. E., 69
Cockcroft, J. D., 690
Compton, A. H., 205
Coulomb, 35
Crookes, W., 99
Curie, P., 146

DALTON, J., 13
Davison, 178
Davy, H., 91
Dirac, P. A. M., 157, 170
Donovan, G. E., 690
Drude, P., 149

EINSTEIN, A., 189

FERMI, E., 157, 170
Fowler, R., 41
Franz, R., 151

GILL, J. F. 425
Goldstein, 99
Graetz, 455

HARTSHORN, S., 541
Hirschlaff, E., 272
Houstoun, R. A., 258

KAUFMANN, W., 102
Korst, D. W., 690
King, A. H., 397

LANGMUIR, I., 276
Laplace, P. S., 112
Latham, R., 397
Lipson, H., 226
Loeb, L. B., 56, 81, 84, 98
Lorentz, H. A., 149
Low, K. S., 222
Ludwig, 433
Lummer, O., 25
Lyman, T., 41

MEEK, J. M., 81, 84, 98
Millikan, R., 110, 111, 188
Morton, C., 690
Mullins, L., 226

NEWMAN, F. H., 142
Nottingham, 92

OLIPHANT, 194

PASCHEN, F., 39, 66, 68
Patchett, G. N., 690
Pauli, W., 52
Peek, 75
Perrin, J., 99
Petry, 192
Pickering, E. C., 41
Planck, M., 24, 27, 30
Polgreen, G. R., 541
Poynting, F. H., 56
Preston, G. D., 258
Preston, T., 28, 56
Pringsheim, E., 25
Pullin, V. E. 226
Puckle, O., 569, 571

RAETHER, 81. 82
Randall, J. T., 272
Richardson, O. W., 152
Rontgen, K., 197
Rushforth, L., 397
Rutherford, E., 34
Rydberg, J. R., 41

SCHIFF, L. T., 690
Schottky, 184
Searle, V. H. L., 142
Slepian, 91, 433
Smith, C., 56
Spreadbury, F. G., 78, 349, 451, 660, 664

Starling, S. G., 200, 251, 459
Stefan, J., 31
Stewart, 103
Stirling, 164
Stokes G., 198, 262

Teago, P. J., 425
Thomas, H. A., 398
Thompson, J. J., 35, 56, 98, 100, 101, 108, 153, 188, 198
Thomson, G. P., 98
Tolman, 103
Tomlin, G. M., 541
Townsend, J., 80

Van der Waals, J. D., 4

Weideman, G., 151
Wein, W., 26, 27, 30
Weiss, P., 146
Whitehead, 75
Wilkins, M. F. H., 272
Williamson, B., 56
Wilson, C. T. R., 107
Wilson, H. A., 109
Wilson, W., 258, 541